POLYPUBLIE
Polytechnique Montréal

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

Titre: Title:	Architecture complètement convolutive à champ d'activation large pour la segmentation sémantique de la vasculature rétinienne dans les images de fond d'oeil
Auteur: Author:	Gabriel Lepetit-Aimon
Date:	2018
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Lepetit-Aimon, G. (2018). Architecture complètement convolutive à champ d'activation large pour la segmentation sémantique de la vasculature rétinienne dans les images de fond d'oeil [Master's thesis, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/3280/

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/3280/
Directeurs de recherche: Advisors:	Farida Cheriet
Programme: Program:	Génie informatique

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal

https://publications.polymtl.ca

UNIVERSITÉ DE MONTRÉAL

ARCHITECTURE COMPLÈTEMENT CONVOLUTIVE À CHAMP D'ACTIVATION LARGE POUR LA SEGMENTATION SÉMANTIQUE DE LA VASCULATURE RÉTINIENNE DANS LES IMAGES DE FOND D'OEIL

GABRIEL LEPETIT-AIMON
DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
AOÛT 2018

© Gabriel Lepetit-Aimon, 2018.

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

ARCHITECTURE COMPLÈTEMENT CONVOLUTIVE À CHAMP D'ACTIVATION LARGE POUR LA SEGMENTATION SÉMANTIQUE DE LA VASCULATURE RÉTINIENNE DANS LES IMAGES DE FOND D'OEIL

présenté par : LEPETIT-AIMON Gabriel
en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées
a été dûment accepté par le jury d'examen constitué de :

M. GUIBAULT François, Ph. D., président
Mme CHERIET Farida, Ph. D., membre et directrice de recherche
M. DUONG Luc, Ph. D., membre

REMERCIEMENTS

Je souhaite tout d'abord remercier chaleureusement ma directrice de recherche Pr. Farida Cheriet qui m'a guidé tout au long de mes travaux de recherche, qui m'a conseillé lorsque j'avais des interrogations sur la marche à suivre et qui m'a éclairé de sa rigueur scientifique lorsque je rédigeais des articles ou mon mémoire. Plus encore, j'aimerais lui exprimer ma profonde reconnaissance pour m'avoir soutenu et réconforté alors que, les dates limites approchant, le stress se faisait de plus en plus pesant ; et pour m'avoir toujours redonné confiance en moi lorsque, au gré de la maitrise, le doute menaçait de désorienter mes recherches.

Je veux aussi remercier l'assistant de recherche au laboratoire LIV4D : Philippe Debanné, pour avoir relu et corrigé mes écrits dans la langue de Shakespeare et pour m'avoir ainsi évité le ridicule des erreurs grammaticales qui semblent apprécier se nicher dans mes textes. Merci aussi pour sa disponibilité et sa promptitude à résoudre n'importe quel problème logistique inévitable lors d'une maitrise.

Je tient à remercier l'ophtalmologiste de l'hôpital Maisonneuve-Rosemont : Pr. Renaud Duval qui, malgré son emploi du temps particulièrement chargé (comme tout ophtalmologiste) à toujours su se rendre disponible lorsque j'avais besoin de précisions sur le savoir médical de l' ?il ou lorsqu'il fallait vérifier la validité de mes annotations.

Je remercie mon complice d'étude et de recherche Clément Playout, avec qui le travail est toujours efficace et minutieux, la conversation toujours passionnée et passionnante, et la cuisine toujours délicieuse et appréciée ! Je remercie aussi mon professeur devenu camarade Fantin Girard pour nos discussions particulièrement enrichissante sur le traitement d'image, les réseaux de neurones et le libéralisme économique.

Enfin j'adresse un remerciement collectif à tout le laboratoire LIV4D ou règne la convivialité et l'esprit de famille. Je vous retrouverai avec grand plaisir l'année prochaine dans la salle du fond pour y débuter mon doctorat et pour profiter du canapé !

RÉSUMÉ

L'imagerie de fond d'œil permet l'observation non-intrusive des structures anatomiques de la rétine. Ces images sont singulièrement très informatives pour évaluer le risque d'apparition de pathologies oculaires, cardiovasculaires ou cérébrovasculaire, dont le dépistage et le traitement préventif sont des défis majeurs de la médecine contemporaine. Plus particulièrement, les anomalies de la micro-vasculature rétienne sont des symptômes connus pour présager de ces maladies. L'extraction automatique et fiable de cette vasculature est donc une étape importante vers la conception d'un algorithme de diagnostique automatique convoité par les cliniciens.

L'extraction de la vasculature rétinienne nécessite l'exécution de deux opérations : d'une part la segmentation des vaisseaux de la rétine, et d'autre part leur classification entre artérioles et veinules. La revue de littérature sur ces deux tâches révèle que les réseaux de neurones convolutifs sont très souvent utilisés pour effectuer la segmentation des vaisseaux, mais presque toujours absents des méthodes de classification vasculaire. En effet, les méthodes de classification les plus performantes appliquent toutes le même protocole : grâce aux outils de la théorie des graphes, elles reconstruisent l'arbre vasculaire rétinien à partir de sa carte de segmentation. Ensuite, un classificateur rudimentaire établi une première labellisation artère/veine pour chaque pixel appartenant à vaisseau. Enfin, ces labels sont moyennés, corrigés et propagés à travers le graphe de l'arbre vasculaire afin que chacun de ses segments soit étiqueté. Ce protocole imite en fait la démarche des cliniciens lorsqu'ils annotent les images de fond d'œil. En effet, mis-à-part les plus gros vaisseaux, il est souvent difficile, voire impossible, de distinguer une artériole d'une veinule simplement par son apparence. Par conséquent, les cliniciens annotent d'abord les vaisseaux larges émergeant du disque optique en fonction de leur couleur (les veinules sont plus foncées que les artérioles) puis suivent ces vaisseaux à travers les bifurcations et les croisements en propageant les labels vers les terminaisons vasculaires. Pour résoudre les bifurcations et les croisements, les méthodes de classifications vasculaires automatiques reposent en général sur des connaissances a priori de l'anatomie des vaisseaux rétiniens formulées sous forme de règles topologiques. Cependant, ces règles ne peuvent pas tenir compte des anomalies de la vasculature rétinienne, puisque ces dernières en sont précisément les exceptions. Ainsi, parce qu'elles sont particulièrement peu robustes aux anomalies de la vasculature rétinienne, ces méthodes sont mal adaptées pour l'analyse d'image de rétines pathologiques.

Le travail de recherche décrit dans ce mémoire propose un algorithme capable d'une part

d'effectuer simultanément la segmentation et la classification de la vasculature rétinienne et d'autre part d'apprendre sa topologie. Le modèle proposé s'inspire d'un réseau de neurone ayant fait ses preuves dans domaine du traitement d'images médicales : le U-Net. En effet, la capacité de ce réseau de neurones complètement convolutif à extraire des caractéristiques d'intensités et de textures le rend particulièrement performant pour résoudre des problèmes de classification non-binaire par pixel appelée *segmentation sémantique*. Mais, ce modèle est mal adapté à la morphologie fine et allongée des vaisseaux et à la similarité d'apparence entre les artérioles et les veinules. L'application naïve de ce modèle aux images de fond d'œil produit d'ailleurs de nombreuses incohérences topologiques de classification. L'objectif de ce travail est donc de perfectionner l'architecture U-Net pour améliorer ses performances de segmentation sémantique de la vasculature rétinienne à partir d'images de fond d'œil. Plus précisément, il cherche d'une part à améliorer la classification des vaisseaux larges et moyens en rendant possible l'apprentissage de caractéristiques topologiques de large échelle ; et d'autre part à améliorer la classification des vaisseaux les plus fins en propageant les labels des vaisseaux larges vers les terminaisons vasculaires.

L'étude structurelle de l'architecture U-Net révèle que son champ d'activation [1] est plutôt restreint : 125×125 pixels, bien trop limité pour pouvoir apprendre des caractéristiques topologiques larges. Pour augmenter le champ d'activation d'un réseau complètement convolutif la méthode traditionnelle consiste à ajouter un étage de sous-échantillonge/convolutions/sur-échantillonage, mais cette modification du modèle est extrêmement couteuse en mémoire et en temps de calcul (l'inférence du modèle passe de 67 Méga Flops à 389). Ce mémoire propose une nouvelle méthode d'augmentation efficace du champ d'activation pour un coût en ressources faible (4 MFlops). L'étage le plus profond du U-Net (situé entre les branches d'encodage et de décodage) travaille à une résolution 16 fois inférieure à celle du patch d'entrée. En faisant l'hypothèse qu'à cette résolution, les caractéristiques topologiques permettant la classification des vaisseaux moyens sont toujours visibles, on peut imaginer les extraire directement de l'image sous-échantillonnée. En pratique, l'extraction est réalisée par une branche opérant à basse résolution (128×128 pixels au lieu de 2048×2048) et dont le champ d'activation est de 21×21 pixels soit l'équivalent de 336×336 pixels à la résolution initiale. Dans un premier temps, cette branche est pré-entrainée seule à effectuer la segmentation sémantique des images basse résolution. Puis elle est raccordée à l'étage le plus profond du réseau par concaténation de ses 256 caractéristiques à celles issues de la branche d'encodage du U-Net. Ces caractéristiques fournissent ainsi des informations contextuelles au réseau et contribuent à sa prédiction finale via la branche de décodage. Cette nouvelle architecture est appelée Réseau Complètement Convolutif à Champ d'Activation Large (Large Receptive

1. la zone de l'image dans laquelle les pixels sont corrélés pour produire le label d'un unique pixel

Field Fully Convolutional Network, LRFFCN). Ce mémoire propose aussi d'intégrer au U-Net une couche implémentant un Champ Aléatoire Conditionnel (Conditional Random Field, CRF) sous forme d'un réseau convolutif récurrent. Cette couche CRF placée à la fin du réseau est capable de propager les labels vers leur voisinage proche à travers les vaisseaux et ainsi de limiter les aberrations topologiques de classification.

Les validations successives de ces deux améliorations de l'architecture U-Net sur la base de données publique AV-DRIVE démontrent que l'architecture LRFFCN apprend avec succès des caractéristiques topologiques larges qui engendrent un gain de performances en classification de 3.8% par rapport à l'architecture U-Net classique. Comparée à l'état de l'art, cette architecture améliore aussi la qualité de segmentation vasculaire puisque la précision de segmentation atteint 96.1% contre 94% dans un article récent employant un apprentissage adversaire. Les performances de la couche CRF sont cependant plus mitigées : si elle parvient bien à corriger certaines anomalies topologiques locales, elle propage aussi des erreurs de classification. Autrement dit, l'algorithme ne parvient pas toujours à propager les labels des vaisseaux larges vers les terminaisons vasculaires et ne rivalise donc pas encore avec les algorithmes utilisant la théorie des graphes, en tous cas lorsqu'il est validé sur des images faiblement pathologiques.

Il reste que le gain de performance de l'architecture LRFFCN prouve qu'elle augmente efficacement le champ d'activation des réseaux complètement convolutifs pour un coût en ressources réduit. De plus, contrairement aux méthodes de la littérature, les performances de cette méthode s'améliorent à mesure que l'ensemble d'entrainement grandit. Ce travail constitue donc un premier pas prometteur vers la segmentation sémantique du réseau vasculaire rétinien et vers l'étude clinique visant à corréler ses anomalies à des pathologies oculaires, cardiovasculaires et cérébrovasculaires.

ABSTRACT

Retinal fundus imaging allows the non-invasive observation of the retinal anatomical structure. Fundus images and more specifically the study of retinal micro-vasculature anomaly, are known to be informative when estimating risks of retinopathy, cardiovascular and cerebrovascular pathologies. Early diagnosis of those pathologies is the key to reducing their mortality rates and is a challenge of modern medicine (cardiovascular diseases is the second cause of deaths in Canada). Thus, an automatic and reliable extraction of the retinal vasculature tree is a key step towards the conception of automatic screening algorithms wished by clinicians.

Extraction of the retinal vasculature tree consist in two tasks: the segmentation of the vessels and their classification between arteries and veins. Deep neural network are often used for the segmentation task but are almost never used for the classification task. Indeed, for this second task, algorithms usually make an extensive usage of the graph theory to reconstruct the retinal vascular tree from the segmentation map. A simple classifier is then used to compute arteries and veins labels which are averaged, corrected, and propagated along the vascular graph. Actually, this method attempt to mimic clinicians behaviour. Because small arteries and veins are not distinguishable by local features, clinicians start by labelling larger vessels (veins are always darker than arteries) and then propagates those labels towards the vascular endings by following each vessels through its bifurcations and crossing. In order to solve those bifurcations and crossing, automatic vascular classifications usually relies on prior anatomical and structural knowledge of the retinal vasculature which are transcribed into topological rules. However they can't take into account vascular anomalies because they are the exceptions of those rules. Thus, because the are not reliable to vascular anomalies, those methods are not well fitted to perform the retinal vasculature extraction with a view to diagnose cardiovascular or cerebrovascular pathologies.

The work described in this report propose an new method to perform the retinal vasculature extraction from fundus images. Unlike state of the art methods, our algorithm perform the segmentation and the classification tasks jointly, so the latter doesn't relies on the first and thus is not sensible to segmentation mistakes. Also our algorithms learn the vascular topologies instead of using prior-knowledge and thus should be less sensitive to topological anomalies. Our model is inspired from a neural network architecture famous in medical imaging processing: the U-Net. This architecture is well known to efficiently extract local intensity and textural features to perform non-binary pixelwise classification also known se-

mantic segmentation. However this model is not well suited to analyse the thin and extensive shapes of retinal vessels. The naïve use of this model on retinal vasculature extraction gives good segmentation results, but shows numerous classification incoherences. The main objective of our research is to improve the U-Net architecture so it reaches better performances in semantic segmentation of the retinal vasculature from fundus images. More precisely, our work aims to improve large vessels classification through the learning of large-scale topological features, and to improve small vessels classification by allowing the propagation of labels towards vascular endings.

The structural analysis of the U-Net architecture reveal that its Receptive Field [2] is quite small: 125x125 pixels, way to narrow to learn large-scale topological features in a high resolution picture. The traditional method to increase the receptive field of a fully-convolutional neural network is to add a pooling/upsampling stage, however because it requires the doubling of the input patches size during training, this method have a heavy memory and computational cost (the inference would need 389 Mega Flops instead of 67 MFlops). This report propose a new method to increase the receptive field of Fully-Convolutional Neural Network at a low computation and memory costs (in our case the inference only need 4 MFlops more). The most deeper stage of the U-Net (between the encoding and the decoding branch) operate at a resolution 16 times inferior to the input patch resolution. However, assuming that topological features are still visible at this low resolution (128x128 pixels instead of 2048x2048), extracting those features at this resolution would be more efficient computation-wise. We propose to learn large-scale features through a simple convolutional branch operating at a low resolution. Its receptive field is only 21x21 pixels wide but is equivalent to 336x336 pixels at the initial resolution. This low-res branch is first trained alone to perform semantic segmentation on low resolution images. Then its model is incorporated into a U-Net model and its 256 deepest features are concatenated with those computed by the encoding branch of the U-Net. Thus the low-res branch brings large-scale contextual features to the core of the U-Net model. We call this new architecture Large Receptive Field Fully Convolutional Network (LRFFCN). We also propose to add Conditional Random Field (CRF) implemented as a recurrent neural network to the ending of our model. This CRF layer enable the propagation of labels towards their neighbourhood and should lower topological classification aberrations.

The validation of our to improvements of the U-Net architecture on the public database AV-DRIVE show that our LRFFCN architecture efficiently learn large-scale topological features enabling a classification accuracy gain of 3.8% when adding the low-res branch. Compared to the state of the art methods our model also improves the segmentation accuracy to 96.1%

2. The size of the area from the input image used to predict the label of a single pixel in a fully-convolutional neural network.

against 94% in a recent paper using adversarial training. However the CRF layer doesn't perform as well as it should. If it successfully propagate labels along vessels, this propagation is quite blind in the sense that true classification and errors are propagated indistinctively. In other words, the CRF layer doesn't succeed in propagating the classification labels towards the vascular endings. Therefore our methods doesn't match state of the art classification performance of graph-based methods.

Though, the performance gain of the LRFFCN architecture proves that it successfully increases the receptive field of Fully Convolutional Neural Network at a reduced computational cost. Moreover, and unlike state of the art methods, our model performance rises as the training database grows. Furthermore its great segmentation performance and its reasonably good classification performance makes it a good model to pre-annotate images, which then only need quick corrections from clinicians to be added to the training database. For all this reasons the work is a promising step towards automatic semantic segmentation of the retinal vascular network and towards a clinic study correlating retinal vascular anomalies and risks of cardiovascular or cerebrovascular pathologies.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

CNN	Réseau de neurone convolutif (Convolution Neural Network)
cf	*confer* (du latin, se référer à)
CRF	Champs aléatoire conditionnel (Conditional Random Field)
FCNN	Réseau de neurone complètement convolutif (Fully Convolution Neural Network)
FPGA	Circuit Logique Programmable (Field-Programmable Gate Array)
k-NN	k-Plus Proches Voisins (k-Nearest Neighbour)
LDA	Analyse Discriminante Linéaire (Linear Discriminant Analysis)
SVM	Machine à Vecteurs de Support (Support Vector Machine)

LISTE DES ANNEXES

CHAPITRE 1 INTRODUCTION

Commercialisées une première fois par Zeiss en 1926, les caméras d'imagerie du fond de l'œil sont aujourd'hui bien connues des services ophtalmologiques. Le procédé d'acquisition non-intrusif est resté le même depuis 1926 : une simple caméra microscope et un faisceau de lumière, tous deux placés juste devant la pupille et dirigés vers le fond d'œil. Presque un siècle plus tard, les techniques d'acquisitions s'étant améliorées (flash électrique, capteurs numériques...), les caméras fundus sont maintenant capables de prendre des clichés de la membrane rétinienne interne à haute résolution (plus de 5000×5000pixels) et en couleur.

Les ophtalmologistes ont depuis bien longtemps regardé leurs patients dans le noir de la pupille, scrutant leurs rétines pour y déceler des lésions ou des anomalies vasculaires. Certaines pathologies de l'œil peuvent en effet provoquer ces symptômes avant d'altérer l'acuité visuelle du patient. Ainsi la rétinopathie diabétique (première cause de cécité au Canada en 2016) affaiblie la paroi des vaisseaux et entraine des lésions de la rétine (anévrismes, hémorragies, exsudats...) ou des anomalies dans sa vasculature (rétrécissement des artères, augmentation de la tortuosité, pincement de la veine lors d'un croisement entre vaisseaux...). Mais les ophtalmologistes ne sont pas les seuls cliniciens à étudier les images de la rétine.

Dès la fin du 19^{e} Gunn (1898) évoquait le potentiel des anomalies de la micro-vasculature rétinienne pour effectuer un diagnostique de maladies cardiovasculaires. De multiples études furent menées sur ce sujet durant le 20^{e} siècle révélant notamment qu'un patient atteint de rétinopathie a deux fois plus de chance de subir un accident vasculaire cérébral. Et ces études continuent encore aujourd'hui. On peut par exemple citer l'*ARCS* (Atherosclerosis Risk in Communities Study), une étude menée entre 1985 et 2016 sur 4 communautés différentes aux États-Unis (12642 personnes âgées de 51 à 72 ans) qui enquête sur l'étiologie (causes et facteurs) de l'athérosclérose et sur les risques cardiovasculaires qu'elle provoque. Parmi les 745 articles publiés dans le cadre de cette étude, une dizaine envisage l'utilisation d'images de rétine pour le diagnostique préventif de l'athérosclérose. De l'autre coté de l'océan, l'étude *Rotterdam Eye Study*, observe une population de 5674 individus âgés de plus de 55 ans (à Rotterdam) dans l'objectif d'établir un lien entre les symptômes de rétinopathie et les maladies du système vasculaire.

Rétinopathies ou maladies cardiovasculaires et cérébrovasculaires, pour ces trois pathologies, un changement de comportement du patient permet d'éviter des traitements intensifs, couteux et contraignants. Pour cette raison, les cliniciens sont à la recherche de moyens de dépistage de ces pathologies. La simplicité et l'efficacité de l'imagerie de fond d'œil en font l'examen

non-intrusif idéal pour effectuer un dépistage à l'échelle d'une population tout en restant bon marché. Des programmes de dépistage par télémédecine existent d'ailleurs déjà au Canada : des centres d'annotations reçoivent des images de fond d'œil de patients en attente de diagnostique, les font examiner par des experts qui évaluent la sévérité de la maladie et éventuellement aiguillent le patient vers un ophtalmologue. Ces plateformes souhaitent de plus en plus intégrer des algorithmes d'intelligence artificielle pour accélérer le traitement des images et soulager les ophtalmologues.

Pour concevoir ces algorithmes de diagnostique automatique, deux approches existent. La première consiste à confier l'intégralité du diagnostique à un modèle entrainé à prédire la gravité d'une maladie à partir d'une telle image. Le modèle sera alors parfaitement optimisé pour le diagnostique de cette maladie mais étendre ses capacités à de nouvelles pathologies nécessitera un nouvel entrainement à partir de zéro. De plus, les modèles d'apprentissage machine, une fois entrainés, se comportent comme des boites noires où il est difficile de comprendre le cheminement ayant conduit à la sélection d'un diagnostique plutôt qu'un autre. Or, dans le domaine médical, la justification du diagnostique est parfois tout aussi importante que le diagnostique en lui-même. Surtout si, après examen, l'algorithme décide d'orienter le patient vers un ophtalmologiste : ce dernier aura besoin des symptômes qui ont déterminés la décision. La seconde approche est justement construite autour de la détection de ces symptômes. Des modèles de perceptions sont entrainés à reconnaitre les structures anatomiques de la rétine et leurs anomalies. Chaque type d'anomalies constitue alors une caractéristique dont l'ensemble permet d'établir un diagnostique selon une procédure établie par les cliniciens (et non plus apprises). Il est ainsi aisé de remonter aux symptômes ou d'étendre la procédure de diagnostique à de nouvelles pathologies.

Dans le cas de l'interprétation automatique d'images de fond d'œil, cette procédure devra s'appuyer notamment sur l'extraction du réseau vasculaire rétinien : plus précisément sur la segmentation des vaisseaux et leur classification entre artérioles et veinules. Durant les trois dernières décennies de nombreux algorithmes furent développés pour réaliser ces taches. D'abord par des approches non-supervisées, puis par des modèles d'apprentissage automatique après que des base de données annotées ait été rendues publiques. L'explosion des performances des approches par réseau de neurones convolutifs a récemment atteint le domaine de la vision par ordinateur pour applications bio-médicales. Ces modèles possèdent déjà les records de qualité de segmentation des vaisseaux de la rétine mais ont cependant des difficultés à effectuer leur classification. À l'heure actuelle, les meilleurs algorithmes de classifications vasculaire sont basés sur la théorie des graphes et parviennent à pleinement exploiter les connaissances topologiques de la vasculature rétinienne. En effet, les vaisseaux les plus larges mis à part, il est particulièrement difficile et peu fiable de classifier un vaisseau uniquement

par son apparence. En réalité, lors de l'interprétation de telles images les cliniciens identifient la classe des plus gros vaisseaux puis propagent ces labels vers les vaisseaux les plus fins. Cette démarche, correctement imitée par les algorithmes utilisant la théorie des graphes, ne semble pas être apprise par les réseaux de neurones convolutifs entrainés à la classification d'artérioles et de veinules.

L'objectif principal de ce projet de recherche consiste donc à améliorer les performances des méthodes de segmentation et de classification automatique du réseau vasculaire rétinien à partir d'images de fond d'œil haute résolution en vue de l'intégration à un algorithme de diagnostique automatique de rétinopathie. Plus précisément, il vise à identifier les limitations des réseaux de neurones convolutifs responsables de leur mauvaise performance de classification vasculaire, puis à proposer et évaluer une architecture adaptée à l'apprentissage de la topologie vasculaire rétinienne.

Ce mémoire s'organise en 5 chapitres. Cette introduction constitue le premier chapitre. Le deuxième est une revue de littérature, il établi les connaissances nécessaires à la manipulation d'images de fond d'œil et effectue un tour d'horizon des méthodes déjà proposées pour segmenter et classifier la vasculature rétinienne. On verra les limites de ces méthodes et la nécessité du développement d'un modèle capable d'apprendre la topologie vasculaire. Le troisième chapitre expose l'architecture de réseau de neurones convolutifs à laquelle mes travaux de recherche ont aboutis. Sa structure et le protocole d'entrainement y sont détaillés. Le quatrième chapitre évalue et discute des performances de l'architecture proposée sur plusieurs bases de données publiques puis expose ses limitations. Enfin, le cinquième et dernier chapitre effectue la synthèse des travaux réalisés et émet des propositions d'améliorations.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Anatomie et imagerie rétinienne

Concevoir un algorithme d'interprétation d'images de fond d'œil nécessite une connaissance minimale de ces images : d'une part sur les structures anatomiques qu'elles exposent et d'autre part sur leur méthode d'acquisition. Cette section éclairci ces deux points et résume les recherches que j'ai menées dans la littérature médicale avant d'entamer ma réflexion algorithmique.

2.1.1 Structures Anatomiques Rétiniennes

Trois structures anatomiques sont particulièrement notables sur les images rétiniennes : la *macula*, le *disque optique* et le *réseau vasculaire* rétinien (cf. figure 2.1). La macula et le disque optique sont des "points cardinaux de la rétine" et permettent d'identifier l'échelle et l'orientation de l'image. Le réseau vasculaire constitue l'objet d'étude de ce mémoire.

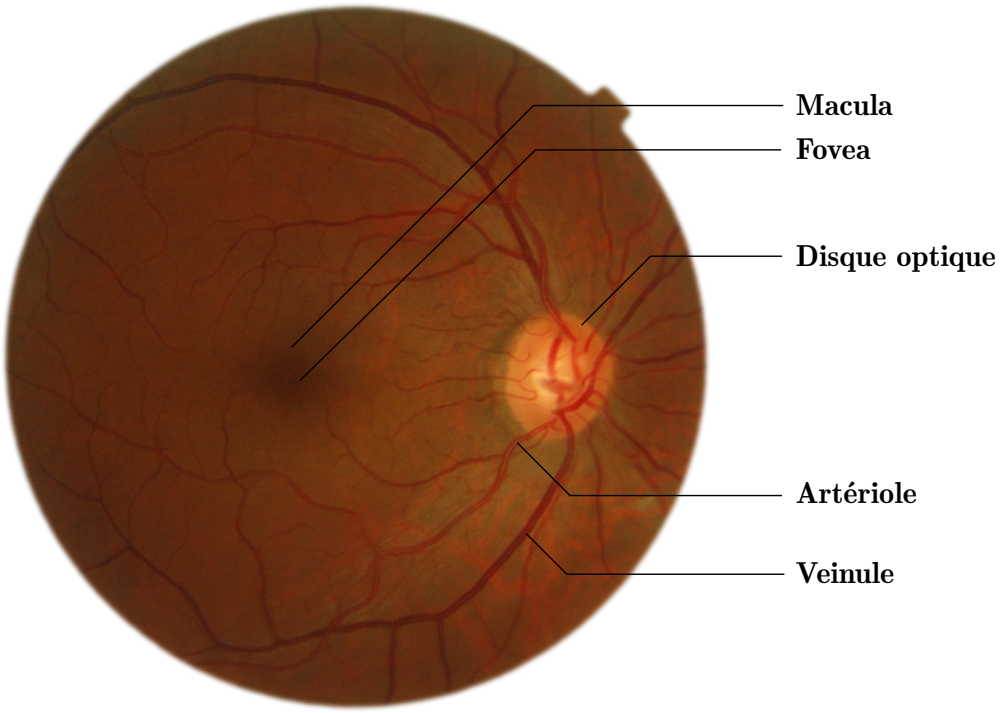

Figure 2.1 Structures anatomiques de la rétine. Crédit : MESSIDOR (Decencière, 2014)

La Macula La captation de l'information lumineuse par la rétine est réalisée par deux types de cellules. Les *cônes* : sensibles aux teintes de couleur sont responsables de la vision diurne (de jour) ; et les *bâtonnets* sensibles au contraste en faible luminosité sont responsables de la vision nocturne. Au centre de la rétine, la densité des cônes augmente fortement, améliorant ainsi l'acuité visuelle au centre du champs de vision. Cette zone, appelée *macula*, capte plus de lumière que le reste de la rétine et apparait donc plus sombre. Son point central, où la densité de cônes est la plus importante, est nommé *fovea* et a un diamètre de 150μm.

Lors d'un diagnostique, les ophtalmologistes portent une attention particulière à la macula. En effet une anomalie vasculaire où une lésion de la rétine porte bien plus atteinte à l'acuité visuelle lorsqu'elle est dans la macula que à la périphérie. Ainsi les acquisitions standards d'image de fond d'œil sont le plus souvent centrées sur la macula. Cependant, dans la perspective de la détection de maladies cardio-vasculaires, toutes les lésions sont importantes, quelque soit leur position. Il faut donc garder à l'esprit lors de la manipulation des bases de données publiques, que les annotations sont susceptibles de biais puisqu'elles sont réalisées par des ophtalmologistes et non par des cardiologues.

Le disque optique Le nerd optique relie chaque globe oculaire à son orbite et permet la liaison nerveuse entre les capteurs de la rétine et le cerveau, il est accompagné par l'artère et les veines ophtalmiques. Le *disque optique* désigne la tête du nerf optique, c'est-à-dire la section du nerf entourée de la papille où converge les axones (terminaisons nerveuses) rétiniens. Le disque optique est situé dans la région nasale de la rétine (à droite de la macula pour un œil droit et à gauche pour un œil-gauche). Il est généralement plus clair que le reste de la rétine car il n'est pas recouvert de capteurs photosensibles. Il existe donc une tache aveugle dans le champs de vision. Le disque optique est la racine de l'arbre vasculaire rétinien : la veine centrale s'y divise en veinules et l'artère centrale (située en nasal de la veine centrale) s'y divise en artérioles.

Le réseau vasculaire L'anatomie de l'arbre vasculaire rétinien varie d'un individu à l'autre et le nombre d'artérioles et de veinules émergeant du disque optique n'est pas fixe. Il existe cependant quelques invariants. La macula est toujours encadrée par deux artérioles (une dans l'hémisphère supérieur et une dans l'hémisphère inférieur). En règle générale, le type de vaisseaux alterne entre artérioles et veinules et deux vaisseaux du même type se croisent très rarement. En cas de croisement, l'artériole sera le plus souvent au dessus de la veinule. Enfin, des vaisseaux très fins émergent parfois du disque optique vers la macula (ils sont visibles sur la figure 2.1), et sont qualifiés d'*artères ciliorétiniennes* (il n'existe pas de veines ciliorétiniennes).

Les artérioles alimentent la rétine en sang chargé en oxygène. Des muscles lisses sont présents dans leurs parois. Cette caractéristique est commune aux artérioles rétiniennes et aux artérioles cérébrales, et elle rend possible le contrôle du débit sanguin qui les traverse. Ainsi l'afflux de sang dans les artérioles de la rétine peut être adaptés en fonction des besoins des cellules photo-sensibles notamment en cas de variation de la luminosité. Les artérioles se dupliquent 3 ou 4 fois avant de laisser la place à des artérioles pré-capillaires. Le calibre des artérioles décroit donc progressivement à chaque bifurcation (où le flux sanguin se divise presque équitablement). Cette règle n'est plus valable au centre de la rétine où les pré-capillaires émergent directement des artérioles principales, des changements rapides de calibre sont alors observables.

Les capillaires tapissent l'intégralité du fond de la rétine (à l'exception de la fovéa, cf. figure 3.1) et permettent, entre autre, l'approvisionnement en oxygène et la récupération du dioxyde de carbone résultant de la respiration cellulaire. Plus précisément, ces échanges entre le sang et les cellules sont permis par l'endothélium, situé sur la paroi des capillaires. Les technologies d'acquisitions non-intrusives ne disposent pas, actuellement, d'une résolution suffisante pour permettre l'observation de ces vaisseaux extrêmement fins ($10 - 15\mu$m).

Les veinules recueillent le sang appauvri en oxygène provenant des capillaires, leur teinte est donc plus sombre que celle des artérioles [1]. On peut aussi distinguer les veinules des artérioles par leurs différences morphologiques : contrairement à ces dernières, la paroi des veinules n'est pas musculaire et possèdent donc des caractéristiques physique différentes (notamment leur réflectivité). De plus, elles sont en moyennes plus larges : 300μm contre 200μm pour les artérioles. Enfin, elles sont généralement plus sujet à la tortuosité que les artérioles.

Ainsi, pour distinguer artérioles et veinules, les cliniciens s'appuient sur des caractéristiques colorimétrique, morphologique ou topologique. Cependant, comme souvent en traitement d'images médicales, les règles énoncées plus haut ne décrivent que les comportements les plus probables. L'importante variabilité entre deux rétines, voire entre deux acquisitions causent de nombreuses exceptions à ces règles. En effet, les méthodes d'acquisitions d'image de fond d'œil n'étant pas standardisées, les teintes et l'illumination des images à traiter fluctuent (figure 2.2) compliquant la tache des algorithmes. De plus, s'appuyer sur des connaissances cliniques a priori pour construire le traitement peut causer des erreurs puisque certaines rétines s'écartent des règles "les plus probables". Par exemple, les croisements de vaisseaux sanguins sont très majoritairement artériole/veinule, pourtant il existe des croisements de vaisseaux du même type, voire des croisements de vaisseaux avec eux-mêmes (cf figure 2.3) !

1. Pour transporter le dioxygène, la protéine d'hémoglobine oxyde les atomes de fer qu'elle contient. L'oxydation du fer résulte en une couleur rouge vif (dans le sang comme pour la rouille). Les artérioles étant plus chargées en oxygènes que les veinules, elles apparaissent naturellement plus claires.

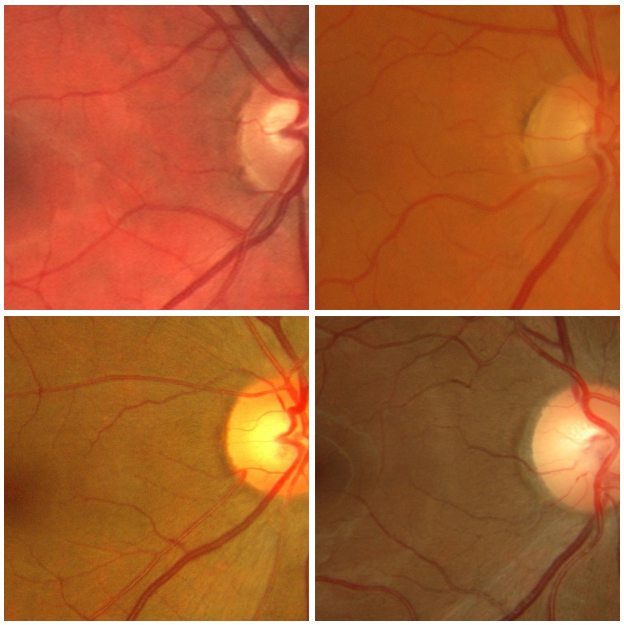

Figure 2.2 Variabilité de couleur de rétine. Les couleurs des images n'ont subi aucun traitement. Crédit : MESSIDOR (Decencière, 2014).

Figure 2.3 Anomalie topologique : boucle. L'image a subit un pré-traitement pour mettre en évidence cette anomalie (hausse de la saturation). Crédit : OPHTA (Decencière et al., 2013).

2.1.2 Imagerie du fond de l'œil

L'imagerie de fond d'œil (fundus en anglais) désigne tout procédé permettant l'acquisition d'une représentation 2D des tissus semi-transparents rétiniens par réflexion de lumière sur la rétine (Abràmoff et al., 2010). Elle s'étend donc de la simple *photographie fundus* monochrome, couleur ou hyper-spectrale jusqu'aux méthodes d'ophtalmoscopie laser (ou Scanning Laser Ophtalmoscopy, SLO) en passant par les angiographies par fluorescence. Pour ce mémoire, nous nous concentrerons seulement sur les photographies fundus couleur.

Images Fundus

Le protocole d'acquisition des images fundus est excessivement simple et est similaire au fonctionnement d'un ophtalmoscope manuel. D'une part une source de lumière blanche et un système optique composé de plusieurs lentilles produisent et concentrent un faisceau lumineux toroïdal vers le fond de l'œil observé. D'autre part, un microscope couplé d'une caméra placé devant l'œil du patient focalisent, agrandissent et effectuent l'acquisition du fond de l'œil au travers de son cristallin et de sa pupille. On obtient alors une image couleur RGB du centre de la rétine (proche de la macula) avec un angle de vue entre 30° et 50° et un zoom $\times 2,5$.

À l'origine les caméras fundus nécessitaient la dilatation de la pupille (par injection d'agent mydriatique). En effet, lorsque la rétine est éclairée, la pupille se rétracte pour empêcher une trop forte intensité lumineuse d'atteindre les capteurs rétiniens ce qui causerait un aveuglement dans le meilleur des cas et une détérioration des cellules photosensibles dans le pire. Dans le cas de l'imagerie rétinienne, ce réflexe métabolique limite considérablement l'exposition du fond de l'œil et dégrade la qualité de l'image. Cependant, l'amélioration de la sensibilité des capteurs des caméras fundus permet aujourd'hui l'acquisition d'images de bonne qualité à travers une pupille rétractée.

La simplicité du protocole d'acquisition d'image Fundus permit une importante diffusion dans les services ophtalmologistes, mais entraina aussi une grande diversité dans les gammes de caméras de fond d'œil. Comme aucun standard n'a été choisi, les images qu'elles capturent possèdent une importante variabilité de luminosité et de balance de couleur auxquelles viennent s'ajouter les aléas de l'acquisition. Il n'est pas rare d'avoir à manipuler des images pas parfaitement nettes (parce que le sujet a bougé ou cligné de l'œil à l'instant où était pris le cliché) ou qu'une poussière soit présente sur l'objectif. Ainsi si les traitements automatiques d'images de fond d'œil sont une voie prometteuse pour le diagnostique préventif, le défi de leur conception réside notamment dans leur robustesse aux variations de qualité des images.

Un nombre relativement conséquent de base de données publiques existent mais peu sont annotées. Les informations relatives aux bases de données publiques les plus importantes ont été reportés dans le tableau 2.1. Ce tableau permet aussi de constater l'amélioration progressive de la résolution des images fundus au fil des années (les entrées sont triées chronologiquement). Généralement les algorithmes de segmentation et de classification du réseau vasculaire rétinien sont comparés en fonction de leurs résultats sur DRIVE (Staal et al., 2004).

Images Fundus Grand Angle (Wide Field)

L'angle réduit (maximum 50°) des images obtenues par caméras fundus classiques n'est pas dû à l'angle de prise de vue de la caméra mais à l'angle d'éclairage. En effet, le faisceau lumineux éclairant le fond de l'œil pour ces caméras étant essentiellement cylindrique et son diamètre étant limité par l'ouverture de la pupille, la surface éclairée de la rétine peut difficilement dépasser 50°, à plus forte raison lorsque la pupille n'est pas dilatée.

Pour palier à cette limitation, Toslak et al. (2017) proposent de ne pas éclairer directement le fond de l'œil à travers la pupille mais plutôt d'utiliser un éclairage indirecte trans-palpébral (au travers de la paupière). En accolant une forte source de lumière blanche LED à la paupière supérieure, ils parvinrent à éclairer suffisamment l'ensemble de la paroi rétinienne

Tableau 2.1 Comparaisons des principales bases de données publiques d'images de fond d'œil

Nom (référence)	Annotations	Taille	Résolution
STARE (Hoover et al., 2000)	— Segmentation des vaisseaux — Classification A/V ponctuelle	402	700 × 650px
DRIVE (Staal et al., 2004) AV-DRIVE (Qureshi et al., 2013)	— Segmentation des vaisseaux — Classification A/V (ligne centrale) — Classification A/V	40	565 × 594px
DIARETDB (T.Kauppi et al., 2007)	— Description textuelle des lésions — Annotations des lésions	130+89	1500 × 1152px
INSPIRE-AVR (Niemeijer et al., 2011)	— Ratio diamètre A/V moyen	40	2392 × 2048px
HRF (Budai et al., 2013)	— Segmentation des vaisseaux	45	5184 × 3456px
MESSIDOR (Decencière, 2014)	— Grade rétinopathique	1200	2240 × 1488px

pour capturer son image sous un angle de 152° sans dilater la pupille. Cette méthode fût qualifiée d'imagerie Grand Angle (Wide Field en anglais). Cette fois l'angle est bien délimité par la bordure de l'iris, qu'on peut d'ailleurs apercevoir sur les bordures des images grand angle (cf. figure 2.4).

Bien que encore peu développée, cette technologie présente un intérêt majeur par rapport aux caméra fundus traditionnelles pour l'analyse du réseau vasculaire de la rétine. D'une part, les vaisseaux plus fins de la périphérie rétinienne sont plus susceptibles de subir des anomalies présageant de maladies vasculaires. D'autre part, les vaisseaux ne sont pas occultés par les limites de l'image ce qui autorise une reconstitution topologique bien plus complète du réseau vasculaire de la rétine.

Cette technologie étant très récente, elle n'a pas encore donné naissance à des bases de données publiques. Mais leur apparitions devraient être imminentes étant donné la fréquence d'installation de caméras grand-angle dans le monde. Lors de la conception d'algorithmes de segmentation et surtout de classification des vaisseaux rétiniens, cette perspective est à garder en tête.

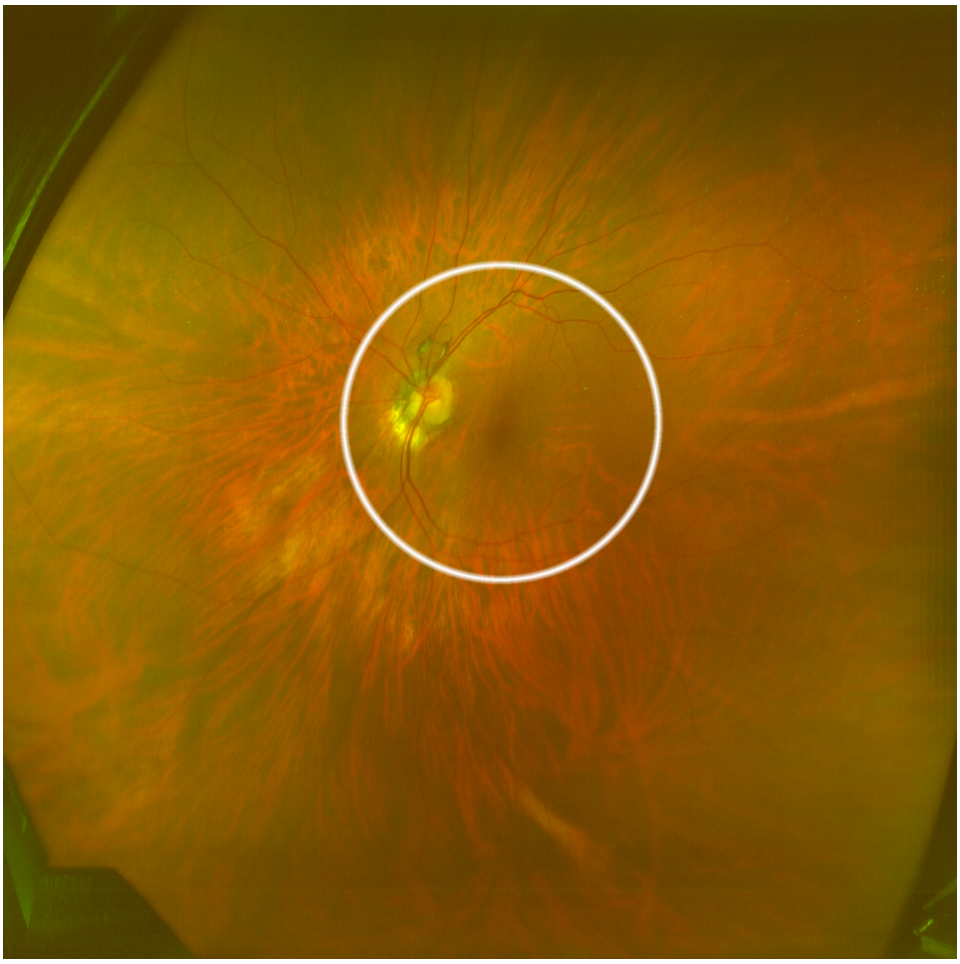

Figure 2.4 Image Grand Angle de la rétine. Le cercle blanc délimite la région capturée avec une caméra fundus traditionnelle.

2.2 Segmentation des vaisseaux rétiniens

Le problème de la segmentation automatique du réseau vasculaire rétinien fût traité dans plus d'une centaine d'article depuis 20 ans. D'abord par des méthodes non-supervisées, puis l'apparition de base de données publiques permirent le développement d'algorithmes supervisés. Enfin, plus récemment, l'explosion des performances des algorithmes d'apprentissage profond est à l'origine de plusieurs modèles de segmentation par réseaux de neurones.

2.2.1 Méthodes non-supervisées

Les méthodes non-supervisées s'appuient sur des connaissances cliniques pour détecter les larges et moyens vaisseaux de la rétine. Elles tirent en général profit de la linéarité des vaisseaux ou de leur contraste avec le fond de l'œil.

Chaudhuri et al. (1989) remarquent que la section des vaisseaux rétiniens peut-être modélisée par une gaussienne et proposent de les détecter en convoluant l'image par des filtres conçus pour imiter le profil de leur section. Les pixels ayant une forte réponse à ces filtres sont sélectionnés comme appartenant à un vaisseau. Cette méthode, appelée *Matched Response Filter* (MRF), fut reprise de nombreuses fois. Hoover et al. (2000) l'améliore en y ajoutant une analyse de caractéristiques à plus large échelle (comme l'histogramme) permettant une meilleure détection des vaisseaux. Meng et al. (2015) perfectionne les filtres en y ajoutant des filtres de Gabor circulaires et des filtres Gaussien multi-échelle, multi-direction.

Chutatape et al. (1998) proposent une méthode effectuant le suivi des vaisseaux (*vessels tracking*). En combinant les filtres décrits plus haut avec des filtres de Kalman, l'algorithme prédit lequel des pixels voisins appartient, avec la plus forte probabilité, à la ligne médiane du vaisseau. Les méthodes qui implémentent le suivi des vaisseaux produisent des résultats topologiquement cohérents mais elles sont aussi extrêmement sensibles aux discontinuités des vaisseaux, qui sont parfois causées par des artéfacts d'acquisitions ou des anomalies vasculaires (par exemple une hémorragie).

Zana and Klein (2001) décrivent un algorithme s'appuyant sur des opérations morphologiques et sur une analyse de courbure. Une ouverture avec des éléments structurants linéaires (suivant différentes orientations) est appliquée à l'image binarisée et une suppression de bruit par opérations morphologiques est effectuée. La segmentation est finalement raffinée par une analyse de la courbature des vaisseaux (calculée via l'opérateur Laplacien). Les structures détectées par cette méthode sont bien linéaires et contiennent peu de faux-négatifs, cependant elles ne sont pas toujours connectées entre-elles.

Dans un article qui décrit un algorithme de détection automatique de rétinopathie hypertensive, K. Noronha (2012) proposent une méthode de segmentation du réseau vasculaire utilisant la transformée de Radon. Appliquée sur la carte de réponse MRF, cette transformée permet d'identifier les structures linéaires dans l'image de fond d'œil. Elle élimine ainsi les artéfacts de la carte MRF et complète les éventuelles discontinuités au sein d'un vaisseau.

Bendaoudi et al. (2016) implémentent une segmentation des vaisseaux rétiniens en temps réel sur des circuits logiques programmables (FPGA) grâce à un algorithme de détection de ligne multi-échelles (Multi-Scale Line Detector, MSLD). Cet algorithme avait initialement été proposé par Nguyen et al. (2013), qui l'appliquaient uniquement sur le canal vert de l'image de fond d'œil. Son objectif était de trouver un bon compromis entre la précision de segmentation et la rapidité d'exécution.

2.2.2 Méthodes supervisées traditionnelles

Les méthodes supervisées segmentent le réseau vasculaire en associant à chaque pixel le label de *vaisseaux* ou d'*arrière-plan*. Ces méthodes utilisent une base de données d'apprentissage pour entrainer leur classificateur à discriminer chaque pixel selon le jeu de caractéristiques qui lui sont associés. Ces méthodes sont en général plus efficaces que les méthodes non-supervisées mais nécessitent la labellisation manuelle du réseau vasculaire sur un nombre suffisant d'images de fond d'œil. Leur développement est donc naturellement plus tardif que les méthodes non-supervisées.

Ainsi, Niemeijer et al. (2004) mènent une étude comparative de méthodes non-supervisées et de leur nouvelle approche où chaque pixel est classifié par un algorithme d'apprentissage automatique. Ils testent 3 classificateurs : k-Plus Proches Voisins (k-Nearest Neighbour, k-NN), un classificateur linéaire et un classificateur quadratique, et concluent sur toutes leur expériences que le k-NN est le plus performant (avec k=30). Le classificateur discrimine les pixels dans un espace à 31 dimensions issues des 31 filtres Gaussien et de leurs dérivées jusqu'au second ordre à 5 échelles différentes. Leur méthode obtient alors de meilleures performances que les autres approches utilisées à l'époque.

Staal et al. (2004) proposent une méthode utilisant aussi un classificateur k-NN mais tirent profit de la topologie des vaisseaux : une analyse du gradient de l'image permet d'obtenir ses lignes de crête (ridge). Ces lignes permettent de détecter les pixels appartenant potentiellement aux lignes médianes des vaisseaux. La segmentation est alors réalisée sur des patchs autour de ces lignes par le classificateur k-NN sur 27 caractéristiques choisies par sélection séquentielle et s'appuyant sur 40 images annotées par leur soins.

Ricci and Perfetti (2007) proposent l'utilisation d'un classificateur plus avancé : une Machine à Vecteur de Support (Support Vector Machine, SVM). Le vecteur de caractéristiques est calculé à partir de détecteurs de lignes et du niveau de gris du pixel classifié. La même année, Ricci et Perfetti rédige un article autre (Perfetti et al., 2007) qui proposent pour la première fois un réseau de neurones cellulaires. Si l'acronyme de ce type d'algorithme est le même que celui des réseaux de neurones convolutifs, l'architecture est assez différentes puisque les neurones des réseaux cellulaires ne sont pas organisés en couche mais peuvent communiquer avec leur voisinage (plus ou moins étendu). Les résultats obtenus sont cependant moins probant que avec les SVMs.

Quatre ans plus tard, You et al. (2011) améliorent les performances de segmentation obtenues par SVM, d'une part en classifiant séparément les vaisseaux larges et les autres, d'autre part en étendant l'apprentissage par une méthode semi-supervisée. Un premier SVM est entrainé

et exécuté sur des images non annotées, les pixels où la confiance dans la prédiction est la plus forte sont retenus, filtrés et ajoutés aux images d'entrainement. Les auteurs ne précisent pas le gain de performance engendré par cette extension de l'ensemble d'entrainement seule (ils obtiennent de meilleurs résultats que les algorithmes précédents). Il reste qu'une approche non-supervisée est attrayante, surtout lorsque l'annotation vasculaire d'une image de fond d'œil peut prendre jusqu'à 2 heures (Staal et al., 2004) ! Le faible nombre d'images d'entrainement est en effet la limitation principale des approches supervisées de segmentation de la vasculature rétinienne.

2.2.3 Méthodes par réseaux de neurones

Enfin, durant les 3 dernières années, plusieurs articles ont appliqué la méthodologie des réseaux de neurones convolutifs à la segmentation des vaisseaux de la rétine.

Maninis et al. (2016) furent parmi les premiers à utiliser les réseaux de neurones convolutifs (Convolution Neural Networks, CNN) pour la détection vasculaire rétinienne. Leur architecture diffère cependant d'un CNN classique : une première branche de 5 couches convolutives (masque 3×3, pas de 2, 2 fois plus de neurones en sortie qu'en entrée) extrait des caractéristiques à différentes échelles, puis ces caractéristiques sont redimensionnées à la résolution de l'image et concaténées. Ce vecteur, contenant donc la sortie de chacune des couches précédentes, alimente une dernière couche convolutive (de masque 1x1) qui effectue leur combinaison linéaire pour prédire la carte vasculaire. En réalité cette architecture fut conçue pour effectuer la segmentation conjointe des vaisseaux et du disque optique, les mêmes caractéristiques issues des couches convolutives sont ainsi combinées d'une part pour détecter les vaisseaux et d'autre part pour détecter le disque optique. Le réseau est entrainé sur ces deux taches en simultané sous un cout de type vraisemblance croisée pondéré pour corriger le déséquilibre des classes. Avec cette architecture, Maninis et al. atteignent les performances de l'état de l'art sans pour autant les dépasser.

L'approche multi-échelle fut aussi favorisée par Fu et al. (2016) avec leur architecture : Deep Vessels. Cette fois la branche de convolutions successives est composée de 4 étages convolutifs chacun étant composé de plusieurs couches convolutives (2 ou 3) et d'une couche de *regroupement* (max-pooling). Chaque étage double le nombre de neurones de sorties (64, 128, 256, 512) mais divise par deux la résolution de ces sorties (par les couches Max Pooling), c'est-à-dire que chaque étage opère à une échelle supérieure que l'étage précédent. Les caractéristiques issues de ces 4 étages sont redimensionnées à la résolution de l'image initiale et corrélée par une couche exploitant les champs aléatoires conditionnels (Conditional Random Fields, CRF) pour prédire une carte de segmentation. Les CRF modélisent le label

d'un pixel par un champ aléatoire de Markov, le reliant à l'intensité des caractéristiques de son voisinage. Contrairement à une simple combinaison linéaire, cette couche CRF corrèle le voisinage du pixel issu de l'image initiale et des caractéristiques et reconstitue une segmentation topologiquement cohérente. Cette approche donne de très bons résultats et permet de dépasser le seuil de 95% de précision sur DRIVE (alors que le second observateur n'obtient que 94.70%).

Depuis, plusieurs travaux ont implémenté des architectures similaires autour des réseaux de neurones convolutifs, améliorant graduellement la précision de segmentation. Lahiri et al. (2017) proposent une approche par entrainement adversaire. Deux réseaux sont entrainés simultanément par deux fonctions de cout opposées, l'un génère des images semblables aux images d'entrainement, l'autre tente de discriminer les pixels provenant d'un fond d'œil, d'un vaisseau ou d'une fausse image générée par le premier réseau. À mesure que l'entrainement progresse, le premier réseau produit des images de plus en plus proche de celles de l'ensemble d'entrainement tandis que les capacités de généralisation du second s'améliorent. Cette approche semi-supervisée permet de contourner le manque d'image d'entrainement en effectuant une augmentation de données intelligente. Surtout que, plus qu'aucune autre méthode supervisée, l'apprentissage profond est particulièrement gourmand en données annotées avant d'atteindre ses pleines performances. Cependant, les auteurs de l'article dressent un bilan mitigé sur l'architecture adversaire : elle obtient de meilleurs résultats lorsqu'elle est entrainée sur très peu d'images, mais reste moins performante que les méthodes actuelles de l'état de l'art (même entrainée sur plus d'images).

2.3 Classification des vaisseaux rétiniens

La plupart des algorithmes de classifications des vaisseaux en artérioles et veinules (classification A/V) de l'état de l'art suivent à peu près la même méthodologie. Dans un premier temps, la vasculature rétinienne est segmentée (par une des méthodes décrites précédemment). Puis la carte de segmentation est affinée jusqu'à ne contenir plus que la ligne médiane des vaisseaux et ainsi extraire le squelette vasculaire. Enfin, un label (artérioles ou veinules) est attribué à chaque portion du squelette à l'aide d'un algorithme d'apprentissage automatique. Les travaux dans ce domaine se distinguent les uns des autres dans le choix du classificateur et des caractéristiques pertinentes ainsi que dans la méthode de propagation des labels au travers du squelette vasculaire.

2.3.1 Méthodes par Classificateur à Caractéristiques Locales

Niemeijer et al. (2011) proposent de labelliser chaque pixel des lignes médianes de vaisseaux à partir de caractéristiques de couleur et d'intensité (teinte, saturation, intensité, et les canaux rouge et vert). Ces caractéristiques sont extraites sur le pixel à classifier et le long de la section du vaisseau passant par ce pixel. Après avoir testé plusieurs classificateurs, ils conclurent que l'algorithme k-NN est le plus performant pour déterminer des labels bruts. Le label final est calculé en moyennant les labels des pixels appartenant à une même portion de vaisseaux, de sorte que des pixels connectés aient une classification analogue.

Cette dernière étape introduit une connaissance a priori dans la procédure de classification. Certes, cette connaissance est très limitée (un vaisseau ne peut commencer comme une veine et finir comme une artère, et inversement), mais selon cette même idée, plusieurs méthodes tirent profit des connaissances cliniques de la topologie vasculaire pour améliorer les performances de classification. Ainsi Vázquez et al. (2012) utilisent un algorithme de regroupement pour diviser l'image du fond d'œil en quadrants autour du disque optique. Puisque chaque quadrant contient au moins une veine et une artère (en tous cas dans la plupart des rétines), la racine de chaque vaisseau peut être comparée avec ses voisines au sein du quadrant. L'un d'eux est nécessairement du type opposé, l'identification est alors renforcée par la comparaison des intensités de couleur des vaisseaux voisins et est moins sensible aux variations de luminosité.

2.3.2 Méthodes par Théorie des Graphes

Des recherches plus récentes tentent de corriger les erreurs structurelles de la classification pixel par pixel et d'augmenter sa cohérence topologique en reconstruisant l'arbre vasculaire et ainsi en identifiant chacun des vaisseaux qui le compose. Ces méthodes ont abondamment recours à la théorie des graphes pour suivre les vaisseaux à partir de la carte du squelette vasculaire. Chaque nœud du graphe est interprété comme une intersection entre deux vaisseaux ou comme une bifurcation d'un vaisseau.

Dashtbozorg et al. (2014) proposent un jeu de règles qui analysent les nœuds du squelette en fonction du calibre des vaisseaux, de l'angle d'intersection et de sa distance. L'application de ces règles permet de corriger le graphe vasculaire et d'identifier un ensemble d'arbres indépendants. Plus particulièrement, ces règles éliminent des erreurs issues de la segmentation : lien incomplet, faux lien (faux-positif) et résout les intersections entre vaisseaux. Après ces corrections, chacun des arbres identifiés doit être propre à un vaisseau et peut donc être classifié en entier. Les auteurs choisirent un classificateur par analyse discriminante linéaire

(LDA) à partir de simples caractéristiques d'intensité.

Pellegrini et al. (2018) décrivent une méthode similaire pour traiter des images grand angle (Wide-Field fundus). Ils raffinent la carte de segmentation vasculaire à l'aide de règles topologiques apprises de manière empirique, puis la convertissent en graphe et augmentent ces segments avec un label A/V issu d'une classification locale. Finalement, une classification A/V générale est effectuée par un algorithme Graph-Cut.

Ces deux méthodes ont recours à des règles et des seuils définis manuellement pour effectuer le suivi des vaisseaux et extraire le réseau vasculaire. Cependant, on l'a vu, la vasculature rétinienne est sujette à une forte variabilité topologique. En particulier, les rétinopathies causent des anomalies (variation anormale du calibre d'un vaisseau à cause d'un rétrécissement local, angle d'intersection peu commun à cause d'une tortuosité importante...). Ces règles statiques risquent d'éliminer ces anomalies du graphe qui sont pourtant capitales dans la détection des maladies : ainsi Pellegrini et al. eurent des difficultés à distinguer les bifurcations des intersections lorsque un segment de vaisseau n'a pas été segmenté.

Estrada et al. (2015a) étendent un de leur précédents travaux pour reconstruire l'arbre vasculaire rétinien de manière fiable (ce travail estimait la topologie d'arbres naturels à partir d'image 2D en les projetant dans un espace de plus grande dimension (Estrada et al., 2015b)). Ils explorent ensuite l'ensemble des arbres labellisés : des arbres A/V sont échantillonnés, une vraisemblance est associée à chacun d'entre eux en tenant compte de la luminosité du vaisseau (indication A/V), d'un modèle de croissance locale et du taux de chevauchement de vaisseaux du même type. Enfin, ils raffinent l'arbre le plus probable par une recherche *Best-First* (Best-First Search, BFS). Cette méthode offre des résultats remarquablement bons, mais nécessite que les vaisseaux ne soient pas déconnectés pour reconstruire l'arbre vasculaire. Lors de l'évaluation de leur méthode ils durent d'ailleurs éliminer une image de la base de test DRIVE car trop de vaisseaux étaient déconnectés du disque optique. De plus l'algorithme BFS est lourd en calcul : la méthode a besoin d'environ 2 minutes pour labelliser une image.

Récemment, Huang et al. (2018) notaient que dans les méthodes présentées jusqu'alors, la classification A/V par pixel avait été relativement négligée et qu'elle reposait en général sur des caractéristiques locales de couleurs, trop simples et pas suffisamment fiables. Ils proposent donc d'extraire visuellement les propriétés de réflexion lumineuse des vaisseaux qui constituent une caractéristique physiquement plus sensée pour discriminer les artérioles des veinules. Cependant cette méthode éprouve des difficultés pour labelliser les moyens et petits vaisseaux dont l'apparence est souvent très semblable. Il reste que leur remarque sur l'importance de la sélection de caractéristiques fiables semble légitime.

2.3.3 Méthodes par Réseaux de Neurones Convolutifs

En réalité les cliniciens utilisent des règles topologiques uniquement pour résoudre certains cas où la structure du vaisseau est particulièrement incertaine. Mais dans la majorité des cas, ils s'appuient plutôt sur la capacité du cerveau humain à suivre le tracer d'un vaisseau : à regrouper des primitives visuelles dans une même entité et à remplir les trous pour restituer l'intégrité d'objets partiellement occlus. Il a été prouvé que ces mécanismes de perception, difficiles à décrire par un algorithme, sont efficacement imités par les réseaux de neurones convolutifs. En effet, ces modèles sont capables d'apprendre des caractéristiques locales de couleur, d'intensité ou de texture mais aussi de la corréler avec des informations contextuelles plus larges (si le patch en entrée est suffisamment grand). De plus, contrairement aux méthodes précédentes, le processus d'apprentissage itératif des CNNs leur permet d'améliorer leur performance à mesure que la base d'entrainement grandit. Pourtant, à ma connaissance, peu de travaux ont appliqué les CNNs à la classification des vaisseaux rétiniens.

Welikala et al. (2017) a entrainé un modèle CNN simple (3 couches convolutives, 2 couches de sous-échantillonnage, 3 couches complètement connectées) qui classifient les pixels des lignes médianes des vaisseaux à partir d'un ensemble de patches 25×25 pixels extraits autour du pixel à traiter. Puis la prédiction du CNN est moyennée sur l'ensemble des pixels d'un segment vasculaire. La taille du patch est particulièrement réduite et les caractéristiques extraites par le CNN sont naturellement locales. Pourtant les auteurs disent atteindre les performances de l'état de l'art sur la base de données DRIVE. En réalité, ils ont modifié le découpage initial en ensemble d'entrainement et ensemble de test puisqu'ils ont réduit ce dernier de moitié (10 images au lieu de 20), pour augmenter d'autant l'ensemble d'entrainement. Il est donc difficile de comparer leurs résultats avec les performances de l'état de l'art.

De plus, cette méthode (comme la plupart des méthodes décrites précédemment) ne tente pas d'apprendre la topologie de la vasculature rétinienne. Elle s'appuie plutôt sur la segmentation de la vasculature et des connaissances cliniques pour corriger grossièrement les erreurs de classification par pixel. Cependant, ces erreurs pourraient être évitées si la classification ne reposait pas seulement sur des caractéristiques locales dérivant de l'apparence du vaisseau, ces caractéristiques étant parfaitement insuffisantes pour discriminer les petites artérioles des petites veinules. Pour obtenir des résultats cohérents avec la perception des cliniciens il faudrait que leur démarche puisse être apprise par le réseau : classifier les vaisseaux les plus larges d'après leur apparence puis propager leur label vers les vaisseaux plus fins qui leur sont connectés. Il faut alors que le réseau apprenne aussi des informations topologiques (suivi vasculaire) et contextuelles (un vaisseau a peu de chance de croiser un autre vaisseau du même type).

2.4 Segmentation Sémantique par réseau de neurones convolutifs

Une solution pour éviter que la classification n'ait à s'appuyer sur la segmentation mais qu'elle puisse tout de même reposer sur des caractéristiques topologiques, est de reformuler le problème en terme de segmentation sémantique. Le double problème de classification à 2 classes (segmentation puis classification) devient alors un unique problème de classification à 3 classes (fond, artères ou veines). Une architecture de réseau est particulièrement adaptée à cette tâche : les réseaux complètements convolutifs (Fully Convolutional Neural Network, *FCNN*).

2.4.1 Réseaux de neurones complètement convolutifs

Contrairement, aux CNNs classiques, les FCNNs ne calculent pas un unique label par patch mais tout une carte de segmentation sémantique où chaque pixel se voit associer un label. Pour ce faire l'architecture des réseaux FCNNs sont généralement symétriques : la première moitié du réseau est formée par une branche d'encodage qui fait graduellement décroitre la résolution des cartes mais augmente le nombre de leurs caractéristiques, la seconde branche réalise l'opération inverse, la résolution augmente progressivement pour atteindre la résolution initiale et le nombre de caractéristiques décroit jusqu'à atteindre le nombre de classes. Les sous-échantillonnages (dans la branche d'encodage) sont généralement réalisés par max-pooling ou en ajoutant un pas aux couches convolutives, tandis que les sur-échantillonnages (dans la branche de décodage) sont implémentés par interpolation du plus proche voisins ou par déconvolutions.

Il y a 3 ans, Ronneberger et al. (2015) proposèrent l'architecture *U-Net* : une architecture de FCNN à laquelle ils ajoutèrent des connexions raccourcies entre la branche d'encodage et de décodage afin d'améliorer les performances de segmentation du modèle. En effet, les U-Nets avaient initialement été développés pour la segmentation de cellules, tâche sur laquelle ils dépassèrent les méthodes de l'état de l'art. Depuis, cette architecture a gagné une importante popularité et a prouvé son efficacité sur de nombreuses applications de segmentation d'images médicales. Cependant, l'application naïve de ce modèle à la segmentation sémantique du réseau vasculaire rétinien ne donne pas les résultats attendus. Si la segmentation est bonne, la classification des artères et des veines n'est pas cohérente le long des vaisseaux. Il n'est ainsi pas rare qu'un groupe de pixels soit soudainement classifiés comme appartenant à une veine alors qu'ils sont entourés d'artères (cf figure 2.5). Notre hypothèse est que ces anomalies de classifications sont dues *au champs d'activation* trop réduit du U-Net.

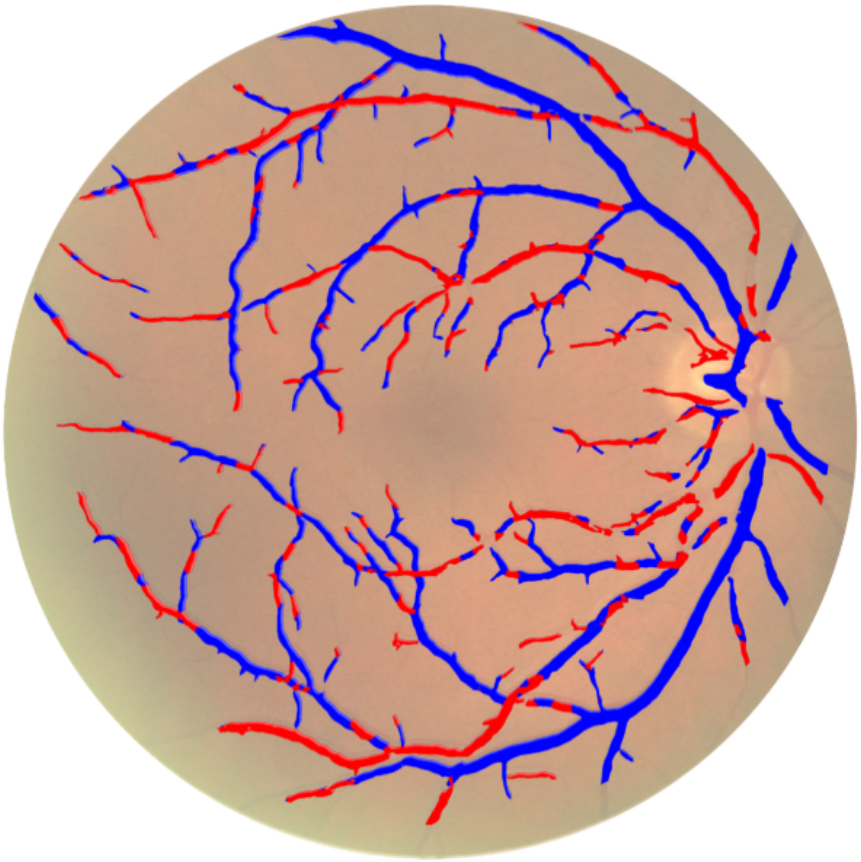

Figure 2.5 Démonstration des mauvaises performances de segmentation sémantique de l'architecture U-Net naïve.

2.4.2 Champs d'activation des réseaux complètement convolutifs

Comme une grande partie du vocabulaire d'apprentissage profond, le *champ d'activation* (en anglais Receptive Field, *RF*) est un terme hérité des neurosciences qui désigne la région de l'espace sensoriel (par exemple le champs de vision) dans laquelle un stimuli entraine l'activation d'un neurone. Dans le cas des réseaux complètement convolutifs, il décrit la taille minimale d'un patch nécessaire à toutes les convolutions du réseau pour calculer le label d'un unique pixel, autrement dit c'est la distance maximale qui sépare deux pixels pouvant être corrélés lors de l'inférence. Ce concept est propre au FCNN, les CNNs étant terminés par des couches complètement connectées leur champ d'activation est égal à la taille du patch d'entrée. Ainsi, bien que les U-Net sont entrainés par des patches de taille 500×500 pixels, le champs d'activation de leur branche d'encodage se limite à 125×125 pixels ! Or, lors de la manipulation d'images de fond d'œil haute résolution (2048×2048 pixels), cette petite fenêtre ne permet pas toujours de distinguer les artères des veines (cf figure 2.6).

Figure 2.6 Exemple de vaisseaux indistinguables avec un champ d'activation limité à 125×125 pixels. (Le vaisseau horizontal est une artériole et celui vertical est une veinule.)

Plusieurs modifications peuvent être opérées sur le modèle pour en augmenter le champs d'activation (RF).

- ajouter une nouvelle couche convolutive avec une taille de masque n augmentera le RF de $n - 1$
- augmenter la taille d'un masque de convolution de n augmentera le RF de n
- ajouter un pas de sous-échantillonnage (stride) de s multipliera le RF par s

En pratique, Luo et al. (2016) ont montré que le Champ d'Activation Efficace (Effective Receptive Field, *ERF*) est toujours plus réduit que le RF théorique. En effet, lors d'une convolution 3×3, le pixel au milieu de l'image intervient dans le calcul de la convolution centrée sur lui et sur celles centrées sur ses 8 voisins. Mais le pixel dans un coin de l'image intervient seulement dans le calcul de la convolution centrée sur un de ses voisins (en mode valide). Autrement dit, un pixel au centre a plus de poids qu'un pixel en bordure de l'image. Lorsqu'on empile de nombreuses couches, on peut même considérer que les pixels en bordures deviennent négligeables. Plus précisément, les auteurs montrent que l'ERF (la carte de contribution) suit une loi gaussienne avec pour origine le centre du patch et dont l'écart type dépend des opérations réalisées dans le réseau. En particulier, l'empilement successif de couches convolutives et les connexions raccourcies réduisent fortement cet écart-type. Au contraire, l'ajout d'une couche de sous-échantillonnage augmente efficacement l'ERF.

2.5 Objectifs de recherche

En résumé, les caractéristiques de couleur et d'intensité, particulièrement efficaces pour détecter les vaisseaux dans une image de fond d'œil, ne permettent pas de distinguer les artérioles des veinules de manière fiable. Certes, la fonction et donc la morphologie de ces vaisseaux sont différentes : les veinules sont en moyenne plus sombres et plus épaisses que les artérioles et ces spécificités sont suffisantes pour labelliser les plus gros vaisseaux. Mais les variations importantes de luminosité et de contraste dues à l'absence d'uniformisation des caméras, les artéfacts d'acquisitions et les anomalies naturelles intrinsèques à toute population d'êtres vivants, gênent la classification des vaisseaux plus fins.

De ce fait, il n'est pas étonnant que les réseaux de neurones convolutifs excellent dans la segmentation vasculaire mais atteignent difficilement les performances de classifications des méthodes de l'état de l'art. Ces méthodes utilisent la théorie des graphes pour reconstruire l'arbre vasculaire rétinien afin de propager la labellisation des vaisseaux les plus larges vers les plus fins. Pour ce faire elles tirent profit de connaissances a priori de la topologie vasculaire rétinienne afin d'interpréter correctement intersections et bifurcations. Ainsi, la fiabilité de ces méthodes reposent d'une part sur la qualité de la segmentation (qui sert à calculer l'arbre vasculaire) et d'autre part sur la conformité des images analysées aux règles anatomiques moyennes sur lesquelles l'algorithme s'appuie. Or, des anomalies topologiques de la vasculature rétinienne ne sont pas rares et sont d'ailleurs souvent symptôme de pathologies. Notre méthode d'extraction de la vasculature doit donc être robuste à ces anomalies pour qu'elle puisse être intégrée dans un algorithme de diagnostique automatique d'image de fond d'œil.

Plutôt que de définir un jeu de règles fixes, l'apprentissage de la topologie des vaisseaux (notamment à partir d'images pathologiques) est la clé pour la segmentation et la classification robuste du réseau vasculaire rétinien. Les réseaux de neurones convolutifs U-Net, reconnus pour leurs excellentes performances en segmentation sémantique, sont une piste de recherche prometteuse mais leur architecture doit-être adaptée.

Ainsi, les objectifs spécifiques de ce travail de recherche sont :

1. Modifier l'architecture du U-Net pour améliorer ses performances en segmentation sémantique de la vasculature rétinienne.
2. Augmenter son champ d'activation pour mieux classifier vaisseaux moyens et larges.
3. Permettre la propagation des labels d'artères/veines à courtes échelles pour améliorer la classification des vaisseaux les plus petits.
4. Évaluer et valider l'architecture sur une base de données publique.

CHAPITRE 3 MÉTHODOLOGIE

Ce chapitre présente la solution proposée pour segmenter et classifier le réseau vasculaire rétinien sur une image de fond d'œil grâce à un réseau de neurones convolutif. Il décrit le prétraitement des images, le choix d'architecture du réseau et sa méthodologie d'entrainement.

3.1 Prétraitement en vue de l'apprentissage profond

Le prétraitement est un sujet presque polémique dans le cadre de travaux d'apprentissage profond. En effet, le principe de l'apprentissage profond est de dériver les représentations successives pertinentes à une tâche. Dès lors, le prétraitement, c'est-à-dire la modification manuelle de la représentation initiale des données, revient à remettre en cause la capacité de l'algorithme à réaliser la tache pour laquelle il a été conçu : extraire les meilleures caractéristiques pour une application donnée. Sans oublier que la conception manuelle d'un prétraitement est conditionnée par notre perception des images de la rétine, or cette perception générale n'est pas forcément la plus adaptée pour une tâche précise et diffère certainement de celle du réseau. Dans le cas de réseaux de neurones convolutifs, tous les pré-traitements reposant sur la convolution de filtres sont donc superflus puisque ces réseaux ont été spécialement conçus pour déterminer les meilleurs filtres convolutifs. La réduction du bruit de la caméra par filtrage gaussien et l'augmentation locale des contrastes par filtrage passe-haut ne seront donc pas imposer, mais pourront être imiter par le réseau si ils sont utiles à la classification finale.

Cependant, les réseaux de neurones complètement convolutifs ne sont pas capables d'apprendre des caractéristiques qui dépendent de l'image entière. Ainsi tout pré-traitement qui unifie une propriété à l'échelle de l'image permet de diminuer la variabilité de l'ensemble d'entrainement, et donc simplifie la tache du réseau et accélère son apprentissage. Lorsqu'on dispose de peu de données pour entrainer le réseau, il est préférable d'avoir de nombreux échantillons décrivant le même mode de l'espace des données d'entrée, plutôt que de nombreux modes présentés par peu d'échantillons.

On l'a vu, les images de fond d'œil présente d'importantes variabilités de luminosité y compris au sein d'un même cliché. L'unification de l'illumination sera donc l'objectif principal de ce prétraitement. Il a été conçu pour être appliqué sur des images de résolutions 2048×2048.

Détection du masque de l'image Les images fundus sont circulaires (à l'exception d'une bosse sur le coin supérieur droit qui permet de distinguer l'œil gauche de l'œil droit). Les contours de l'image sont normalement noirs mais les formats de compression induisent parfois des artéfacts gris foncés dans cette zone. Le masque de l'image M est donc détecté par seuillage du flou médian du canal vert et appliqué à l'image I.

$$M = \mathsf{flouMedian}_{51\times51}(I_{vert}) > 15$$

Unification de l'illumination par filtrage médian Le prétraitement effectue ensuite une première correction d'illumination locale par soustraction de la médiane locale. Pour chaque canal de l'image I, on calcule :

$$I_m = I - \mathsf{flouMedian}_{151\times151}(I) + \mathsf{mediane}(I)$$

Amélioration des contrastes Enfin le prétraitement améliore les constrastes par égalisation d'histogramme adaptatif à contraste limité (Constrast Limited Adaptative Histogram Equalization, CLAHE) sur le canal de luminosité et avec un voisinage de 8×8 pixels. Ainsi les teintes de couleurs ne sont pas modifiées mais les structures anatomiques de la rétine (en particulier les vaisseaux) sont mises en évidence. En notant L le canal de luminosité de l'image dans une représentation colorimétrique TSL (Teinte Saturation Luminosité).

$$L_{clahe} = \mathsf{CLAHE}_{8\times8}(L)$$

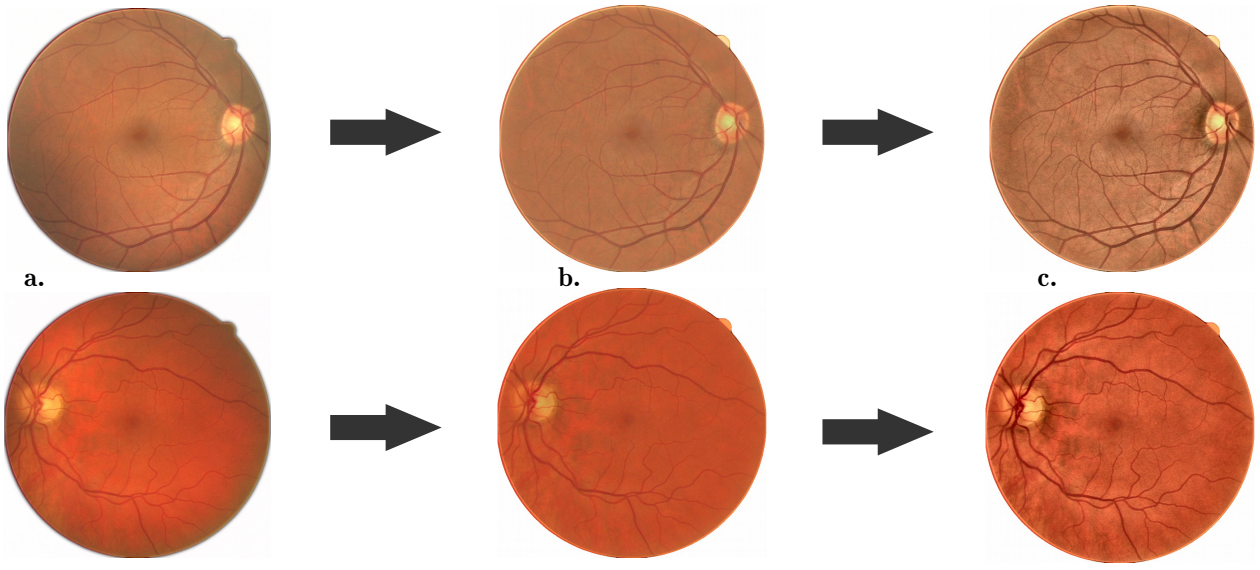

Figure 3.1 Progression d'images après chaque étape du prétraitement. (a. Détection du masque ; b. Filtrage Médian ; c. CLAHE)

3.2 Architecture du réseau de neurones

Le choix d'une architecture de réseau de neurones relève plus d'un processus empirique que d'une science exacte. Bien sûr, la conception d'un modèle neuronal est guidée par quelques règles intuitives : trop de paramètres entraine un risque de sûr-apprentissage (cf. rasoir d'Ockham) ; plus le réseau est profond plus il apprend des représentations abstraites, il est donc préférable d'ajouter des couches convolutives plutôt que d'élargir leur masque ; etc. Mais l'ampleur et l'irrégularité (fortement non convexe) de l'espace des hyper-paramètres doublés d'un temps d'entrainement long rendent impossible une optimisation exhaustive des architectures de réseaux. Ainsi, la majorité des articles s'inspirent fortement de modèles ayant fait leurs preuves, réutilisant parfois jusqu'à leur poids déjà entrainés pour profiter du phénomène d'apprentissage par transfert (transfer learning). Leurs contributions résident souvent dans le choix d'une architecture existante soit en l'appliquant à une nouvelle tâche soit en la modifiant légèrement pour améliorer ses performances. Ce travail s'inscrit dans cette seconde proposition.

3.3 Réseau Complètement Convolutif avec raccourcis

Le modèle conçu pour ce travail s'inspire des U-Nets, très utilisés en segmentation sémantique d'images médicales. Plusieurs points ont motivé le choix de cette architecture.

Elle fait partie de la famille des réseaux complètements connectés (Fully Connected Neural Network, FCNN). Elle est donc composée d'une branche d'encodage CNN classique (couches convolutives et pooling) qui compresse l'information spatiale et permet l'apprentissage de caractéristiques graphiques complexes, suivie d'une branche de décodage qui sur-échantillonne l'information compressée pour obtenir une image de résolution similaire à celle de l'entrée du réseau. Elle possède donc les avantages des réseaux convolutifs (interconnections spatiales et restreintes, redondance des poids, et compression de l'information par pooling), mais est bien plus efficace car l'inférence est réalisée simultanément pour tous les pixels de la carte de segmentation. Elle permet aussi de ne pas fixer la taille du patch en entrée du réseau et ainsi d'entrainer le réseau sur des petits patchs pour augmenter la taille du mini-batch [1], puis de faire l'inférence sur l'image entière.

Contrairement à un CNN classique qui n'aurait été appliqué que sur les pixels détectés comme vaisseaux, l'efficacité de l'inférence des FCNNs permet la *segmentation sémantique*.

1. La taille du mini-batch est le nombre de patches présentés simultanément au réseau lors de son entrainement et sur lesquels les gradients de la fonction de coût seront moyennés. Ainsi à chaque itération de la descente de gradient, la direction de modification des poids est calculé sur plusieurs images et oscille donc moins d'une itération à l'autre.

Au lieu d'effectuer la segmentation des vaisseaux puis leur classifications, ces deux tâches sont réalisées simultanément : chaque pixel étant classifié soit comme une artère, soit comme une veine, soit comme de l'arrière-plan. Les caractéristiques apprises rapidement pour la détection vasculaire (ou plutôt pour la détection de la classe d'arrière-plan, ce qui revient au même) servent de tuteurs pour l'entrainement des caractéristiques de classification. Autrement dit, la classification vasculaire est plus difficile que la segmentation, mais durant l'entrainement, la première peut profiter des caractéristiques apprises pour la seconde.

Enfin, l'architecture U-Net introduit des connections "raccourcis" entre la branche d'encodage et la branche de décodage du FCNN qui permettent un net gain de performances lors de la segmentation d'images haute résolution. En effet, la branche de décodage des FCNN a du mal à restituer seule la précision des formes lors des sur-échantillonnages successifs. Les auteurs du U-Net (Fu et al., 2016) proposent donc de concaténer aux vecteurs de caractéristiques sur-échantillonnées, leurs vecteurs symétriques dans la branche d'encodage juste avant qu'ils soient sous-échantillonnés, c'est-à-dire avant que la précision des formes soit perdue. Les couches convolutives qui suivent ces concaténations combinent donc des caractéristiques précises spatialement mais peu abstraites à des caractéristiques fortement contextualisées mais floues.

L'architecture U-Net n'a évidement pas été conservée telle quelle. L'étage le plus profond du réseau (entre les branches d'encodage et de décodage, noté *étage central* dans la suite du rapport) a été remanié pour intégrer une structure *Fire-Squeeze* proposée par N. Iandola (2016) pour réduire la taille du modèle Alex-Net sans perte de performance. Cette structure éclate les caractéristiques en 3 couches convolutives chacune appliquant un masque de taille différente (1×1, 3×3 et 5×5) puis les recombine par une convolution 1×1 (cf. figure 3.2).

De plus, la taille des patches d'entrainement, initialement 512×512 pixels dans l'architecture U-Net, a été remplacée par 230×230 pixels. Cette réduction par 4 de la quantité de pixels permet d'étendre la taille de mini-batches à 8. À chaque itération, la descente de gradient a ainsi accès à un échantillon plus représentatif de la variabilité de l'espace des images fundus et l'entrainement s'en trouve lissé.

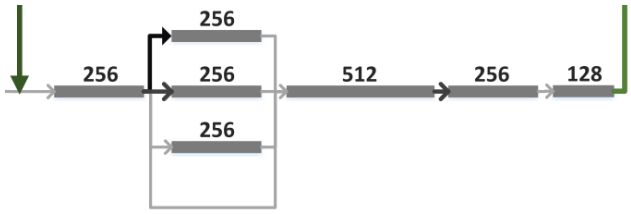

Figure 3.2 Étage central de l'architecture.

Enfin deux modifications majeures ont été implémentées : l'une rend possible l'apprentissage de caractéristiques topologiques à très large échelle et l'autre permet la propagation du label des moyens vaisseaux vers les petits vaisseaux. Les deux prochaines sections sont consacrées à ces modifications. Un schéma de l'architecture complète du réseau est présenté à l'annexe A.

3.4 Branche basse résolution

On l'a vu, le champs d'activation d'un U-Net est de 125×125 pixels, bien trop faible pour permettre l'apprentissage de caractéristiques topologiques à large échelle. Une première solution est d'ajouter un nouvel étage de convolutions, sous-échantillonnage et sur-échantillonnage aux cinq déjà présents. En réalité cette proposition n'est pas viable (en tous cas pas sur la carte NVidia 1070 Ti dont je dispose pour l'entrainement) car le coût en calcul et en mémoire est trop important.

En effet, l'inférence (forward pass) d'un U-Net sur un patch de taille 230×230 nécessite environ 67 Méga Flops. Ajouter un nouvel étage requiert de doubler la taille du patch d'entrée (afin que la taille des cartes de caractéristiques de l'étage central reste inchangée), et le nombre d'opérations nécessaires à l'inférence devient alors 389 MFlops [2], soit une augmentation par un facteur de 5.8 du temps de calcul ! La quantité de mémoire graphique consommée par l'entrainement d'un réseau de neurones est plus complexe à évaluer, mais on peut considérer qu'elle suit la même évolution que le temps de calcul si les sorties de chaque couche sont conservées en mémoire.

L'explosion de complexité décrite à l'instant est entièrement due au dédoublement de la quantité d'informations à l'entrée du réseau. Mais ce surplus n'est pas réellement utile, en réalité il est rapidement estompé par les sous-échantillonnages successifs pour laisser place à des caractéristiques plus précises sémantiquement mais plus floues spatialement. L'intuition guidant ce travail est que ces caractéristiques peuvent-être en grande partie inférées d'une version déjà spatialement floue de l'image : une version redimensionnée à l'échelle de l'étage central du réseau. Ces caractéristiques pourraient alors être calculées séparément directement à partir de l'image sous-échantillonnée et être concaténées au vecteur d'entrée de l'étage central du U-Net. Travailler à une telle échelle permet de considérer une zone beaucoup plus large de l'image : si l'image originale a une résolution 2048×2048 pixels, sa taille équivalente

2. Le nombre d'opérations requises par une inférence a été calculé via les formules théoriques suivantes :
Couche Convolutive : en notant k la taille du masque convolué, (w, h) les dimensions de l'image de sortie, n_{in} et n_{out} le nombre de neurone d'entrée et de sortie, et en posant $K = n_{in} \times k^2$ et $N = n_{out} \times w \times h$, le nombre d'opération est la somme de $O_{convolution} = 2KN$ et $O_{selu} = 2N$.
Sous-échantillonnage par maximum, en utilisant la même nomenclature, le nombre d'opération est $O_{max} = N \times K$

diminue à 128×128 pixels après avoir subit 4 sous-échantillonnages au cœur du réseau, elle peut donc être utilisée dans sa totalité pour calculer des caractéristiques à large-échelle sans craindre un important surcoût calculatoire. L'architecture chargée de l'extraction de ces caractéristiques est aussi inspirée de la structure Fire-Squeeze (cf. figure 3.3) et possède un champ d'activation de 21 pixels de coté, c'est-à-dire, ramené à la résolution initiale : 336 pixels. Le champs d'activation équivalent est donc plus grand que si l'on avait ajouté un étage au U-Net (ce qui l'aurait élevé de 125 à 250 pixels), alors que l'inférence de cette branche basse résolution ne nécessite que 4 MFlops (contre 322 MFlops pour un étage supplémentaire) !

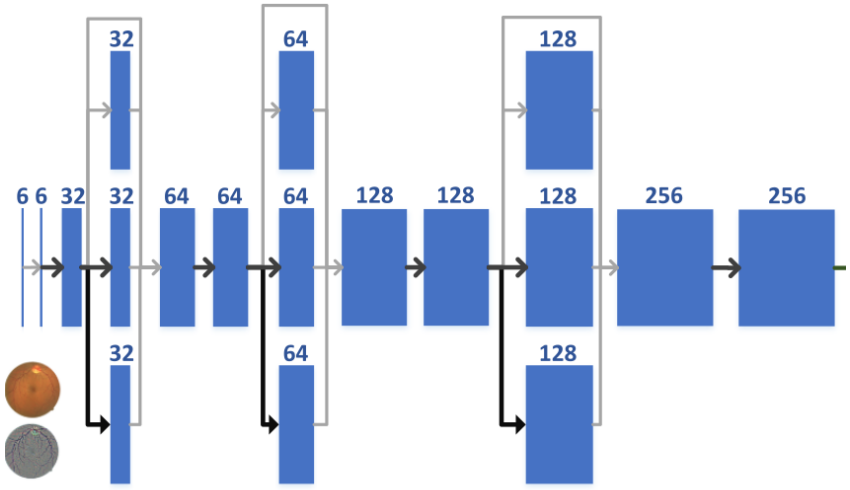

Figure 3.3 Modèle de la branche basse résolution

Pour que la concaténation des caractéristiques provenant du U-Net et celle de la branche basse résolution soit spatialement logique, 3 contraintes doivent-être satisfaites. **1.** L'image à l'entrée de la branche basse résolution doit être sous-échantillonnée avec le même ratio que les caractéristiques du U-Net (en l'occurrence divisée par 16). **2.** Le centre du patch présenté au U-Net doit être le même que celui présenté à la branche basse résolution (à résolution équivalente bien sûr). **3.** Les cartes de caractéristiques en sortie de la branche basse résolution doivent avoir la même résolution que celle en sortie de la branche d'encodage du U-Net afin qu'elles puissent être concaténées. Cette contrainte conditionne la taille du patch en entrée de la branche basse résolution : avec un patch pleine résolution de 230 pixels de coté, la résolution en sortie de la branche d'encodage est de 8 pixels ce qui implique une taille de patch de 28 pixels à l'entrée de la branche basse résolution.

Dans la suite du rapport, l'architecture combinant le U-Net avec cette branche basse résolution sera notée LRFFCN (Large Recepteive Field Fully Convolutional Network, Réseau Complètement Convolutif à Champs d'Activation Large)

3.5 Champs Aléatoires Conditionnels

Lors de l'inférence d'un U-Net (et par extension de tout FCNNs), les pixels sont labellisés en parallèle mais indépendamment les uns des autres. Certes, les pixels voisins partagent certains pixels des cartes de caractéristiques, mais l'approche des réseaux convolutifs attribue un label à un pixel indépendamment du label de ses voisins. À cause de cette approche, les U-Nets souffrent d'incohérences de segmentation quand des classes sont très similaires. Dans notre cas, il n'est pas rare que quelques pixels d'un vaisseau soient classifiés comme une artériole alors qu'ils sont entourés de veinules (cf. figure 3.4).

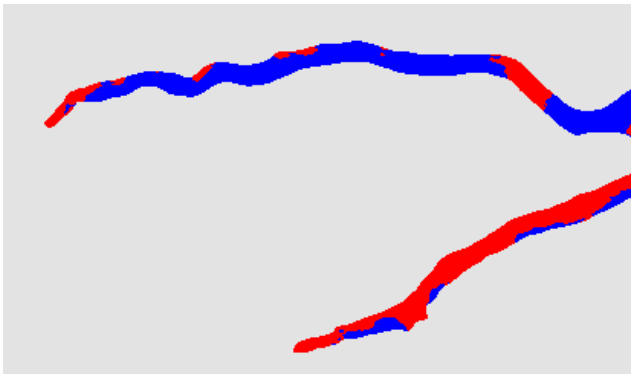

Figure 3.4 Incohérence de segmentation en sortie d'un U-Net. (Code couleur : *rouge* : artèrioles ; *bleu* : veinules ; *gris* : fond).

Ce problème n'est pas propre aux réseaux convolutifs mais concerne tous les algorithmes approchant la classification d'une image pixel par pixel (appelés *classificateurs unitaires*). Afin de corriger leurs erreurs et pour obtenir une segmentation plus cohérente spatialement, Krähenbühl and Koltun (2011) propose une approche probabiliste.

Les explications qui suivent considère le cas simplifié d'une image $\mathbf{I}$ composée de N pixels monochromatiques mais peuvent être aisément étendues aux images RGB. En notant : $\mathbf{I} = \{I_0, I_2, ..., I_N\}$ les intensités des pixels de l'image et $\mathbf{X} = \{X_0, X_1, ..., X_N\}$ les variables aléatoires associées à la labellisation de chacun de ces pixels (avec $X_i \in \{fond, artères, veines\}$). On peut construire un champs aléatoire conditionnel $(\mathbf{I}, \mathbf{X})$, qui associe à une labellisation $\mathbf{x}$ de l'image $\mathbf{I}$, une énergie de Gibbs qu'on cherche à minimiser :

$$E(\mathbf{x}) = \underbrace{\sum_i \psi_u(x_i \mid \mathbf{I})}_{\text{énergie unitaire}} + \underbrace{\sum_{i<j} \psi_p(x_i, x_j \mid \mathbf{I})}_{\text{énergie de parité}} \tag{3.1}$$

Cette énergie est la somme d'une énergie unitaire établie par le classificateur et d'une énergie de parité qui pénalise l'attribution de labels différents à des pixels similaires.

L'énergie de parité dérive d'un potentiel ψ_p défini par :

$$\psi_p(x_i, x_j) = \mu(x_i, x_j)\ k(\mathbf{f}_i, \mathbf{f}_j) \tag{3.2}$$

où μ est une fonction de compatibilité entre les classes des pixels i et j (confondre une artère et une veine est moins pénalisé que de confondre une artère avec le fond) ; et où k est une somme de noyaux gaussiens modélisant la compatibilité des deux pixels selon des caractéristiques arbitraires : $k(\mathbf{f}_i, \mathbf{f}_j) = \sum_{m=1}^{M} w^{(m)} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j)$. Pour la segmentation multi-classes les auteurs proposent d'utiliser les caractéristique suivantes :

$$k(\mathbf{f}_i, \mathbf{f}_j) = \underbrace{w_a \exp\left(-\frac{\Delta_{i,j}^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right)}_{\text{noyau d'apparence}} + \underbrace{w_r \exp\left(-\frac{\Delta_{i,j}^2}{2\theta_\gamma^2}\right)}_{\text{noyau d'uniformité}} \tag{3.3}$$

où I_i et I_j sont l'intensité des pixels i et j ; $\Delta_{i,j}$ est la distance qui les sépare ; θ_α, θ_β et θ_γ sont des coefficients pondérant les caractéristiques de distance et d'intensité ; w_a et w_r sont les poids associés à chacun des noyaux gaussiens. Le premier noyau gaussien retourne une valeur d'autant plus grande que les pixels i et j sont proches et d'apparence similaire, le second n'est sensible qu'à leur proximité. Puisque la distance intervient dans les deux cas en limitant l'amplitude des exponentielles, cette définition des noyaux gaussiens limite l'influence d'un pixel à son voisinage proche (dont l'étendue est réglée par θ_α et θ_γ).

Dans l'article de Krähenbühl and Koltun (2011), l'énergie uniforme était calculée par des classificateurs unitaires mais peut-être remplacée par la prédiction d'un CNN. Zheng et al. (2015) propose une implémentation de cet algorithme par un réseau récurrent de couches convolutives dont la sortie est différentiable par rapport à son entrée. Cette implémentation permet non seulement d'apprendre les paramètres w_a, w_r et $\mu(x_i, x_j)$ mais surtout d'intégrer l'algorithme au sein d'un réseau convolutif et de propager les modifications des poids par la descente de gradient à travers le CRF. Les paramètres des couches convolutives seront donc optimisés en tenant compte des corrections du CRF.

En résumé, l'ajout d'une couche récurrente implémentant les champs aléatoires conditionnels permet de corréler les labels voisins entre eux en utilisant l'image pré-traitée brute comme référence. Autrement dit il permet de propager l'information majoritaire d'un vaisseau pour effacer les anomalies de classification. Les écarts types des gaussiennes ont été réglés à $\theta_\alpha = \theta_\gamma = 25$ pixels, et $\theta_\beta = 10/255$ et l'approximation du CRF est réalisée en 5 récurrences.

3.6 Autres caractéristiques du modèle

Cette section détaille les modifications mineures faites au modèle décrit jusqu'ici.

Fonction d'activation Traditionnellement, on utilise la fonction $\mathrm{ReLU}(x) = \max(0, x)$ pour l'activation non-linéaire des neurones. Par ailleurs, la taille réduite de minibatch (seulement 8 patches) rend l'usage de normalisation par batchs peu efficace. La fonction d'activation a donc été changée pour une fonction SeLU définie par :

$$\mathrm{SeLU}(x) = \lambda \begin{cases} x & \text{si } x > 0 \\ \alpha(e^x - 1) & \text{si } x \leq 0 \end{cases} \tag{3.4}$$

Cette fonction a été proposée par Klambauer et al. (2017) qui ont calculé les valeurs de λ et α pour lesquelles la moyenne et la variance de l'activation des neurones sont liées à un point fixe stable ce qui empêche l'explosion ou la disparition des gradients. Cette fonction d'activation est connue pour remplacer efficacement la normalisation par batchs lorsque celle-ci n'est pas applicable.

Sous-échantillonage mixte Il existe deux méthodes de sous-échantillonnage (pooling layer) particulièrement répandues : par sélection du maximum ou de la moyenne. La première est de loin la plus répandue et s'assimile au fonctionnement des neurones naturels. Yu et al. (2014) propose d'effectuer les deux sous-échantillonnages puis d'en faire la combinaison linéaire avec un coefficient γ appris par la descente de gradient. Ainsi pour chaque caractéristique le sous-échantillonnage est calculé par :

$$P = \gamma P_{maximum} + (1 - \gamma) P_{moyenne} \tag{3.5}$$

3.7 Description de l'entrainement

L'entrainement du réseau est réalisé en 2 étapes. Dans un premier temps la branche basse résolution est entrainée seule sur les images sous-échantillonnées. Puis les poids sont intégrés au réseau complet qui est entrainé sur les images pleine résolution. De manière générale, l'initialisation des poids du réseau est adaptée à la fonction SeLU : ils sont échantillonnés sur une Gaussienne de moyenne nulle et de variance l'inverse du nombre de neurones d'entrée. Une régularisation L2 est appliquée aux masques de convolutions durant l'entrainement qui est guidé par un optimiseur de descente de gradient *Adam*. Enfin, les bases d'entrainements sont augmentées par des transformations géométriques (symétrie horizontale et rotation) et colorimétriques (gamma et contraste).

3.7.1 Entrainement de la branche basse résolution

La branche basse résolution est d'abord entrainée seule en lui ajoutant une couche convolutive 1×1 avec une activation softmax qui combine ses 256 caractéristiques pour obtenir la probabilité d'appartenance aux 3 classes (fond, artère, veine). La branche apprend donc des caractéristiques adéquates à fournir un contexte significatif pour distinguer les vaisseaux. Pour améliorer la qualité de ce contexte, les cartes de labels servant à entrainer la branche ne sont pas simplement une mise à l'échelle des cartes pleines résolutions. Durant le sous-échantillonnage, les vaisseaux sont favorisés sur le fond ce qui a pour effet d'élargir les vaisseaux les plus fins (cf figure 3.5).

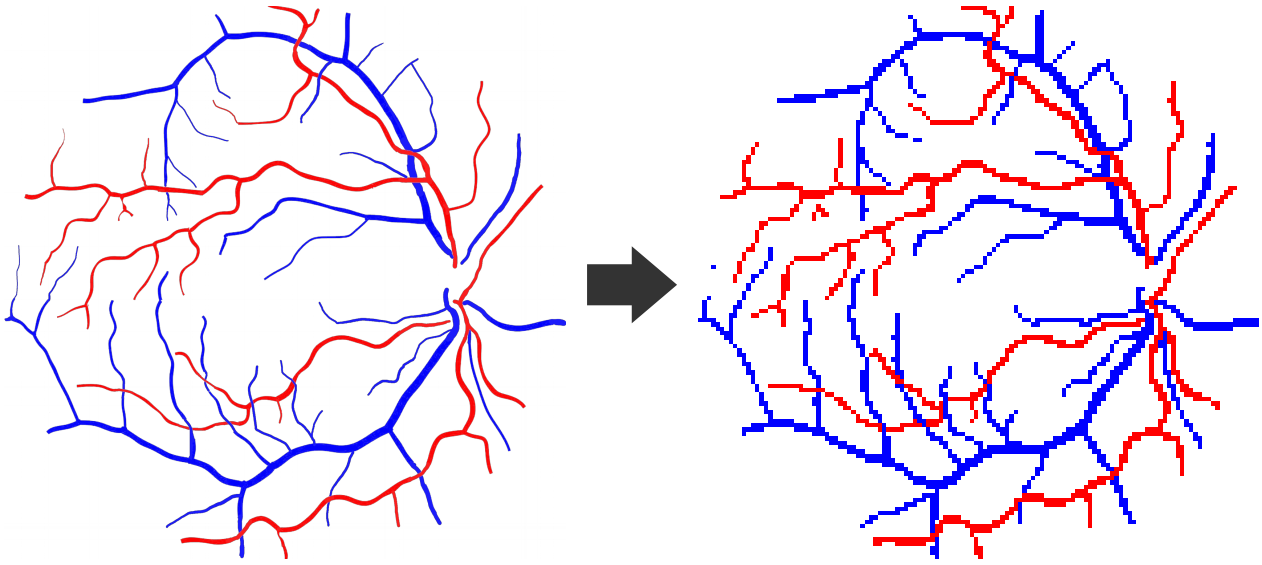

Figure 3.5 Sous-échantillonnage des labels destinés à la branche basse résolution.

La base d'entrainement est composée de 130 images de 128×128 pixels, annotées par mes soins et validées par un expert. L'annotation est réalisé directement sur les images sous-échantillonnés et est donc bien plus rapide (5 minutes plutôt que 2 heures). La majorité de ces images (110 images) sont issues de la base MESSIDOR et les 20 autres proviennent de la base d'entrainement de DRIVE. L'ensemble d'entrainement a été multiplié par 5 par l'augmentation de données décrites plus haut.

La branche basse résolution est entrainée pendant 100 époques avec un coût de type DICE généralisé, très connu pour son efficacité à entrainer des modèle dédié à la segmentation.

En notant $p_n^{(k)}$ la probabilité que le pixel n appartienne à la classe k telle que calculée par le réseau et $q_n^{(k)} \in \{0; 1\}$ la vérité terrain attendue pour ce pixel, le coût DICE généralisé est alors défini par :

$$\mathcal{L}_{\text{DICE}}(P|Q) = 1 - 2\frac{\sum_k \left(\frac{1}{w^{(k)}} \sum_n p_n^{(k)} \times q_n^{(k)} \right)}{\sum_k \left(\frac{1}{w^{(k)}} \sum_n p_n^{(k)} + q_n^{(k)} \right)} \qquad \text{avec } w^{(k)} = \left(\sum_n q_n^{(k)} \right)^2 \tag{3.6}$$

La pondération $w^{(k)}$ permet de corriger le déséquilibre d'occurrence entre les classes.

3.7.2 Entrainement du modèle complet

Dans un second temps, les poids pré-entrainés de la branche basse résolution sont intégrés au réseau principal (sans la dernière couche de classification) et l'ensemble est entrainé sur des images haute-résolution en présentant toujours le patch sous-échantillonné adéquat à la branche basse résolution. Lors de cette étape les poids de cette branche sont aussi entrainés ce qui permet de les adapter à la nouvelle position qu'ils occupent dans le réseau.

Le modèle complet est entrainé sur une base plus réduite de 50 images (30 images de MESSIDOR et 20 images de DRIVE redimensionnées à la résolution de 2048×2048 pixels), annotées selon le même dispositif que pour la branche basse résolution. La taille de l'ensemble d'entrainement est doublé par augmentation de données et, dans chacune des images, 125 patches sont découpés autour du réseau vasculaire. L'entrainement est réalisé sous un coût de type entropie croisée (le coût DICE n'obtenait pas de bons résultats) pendant 30 époques sans le CRF puis pendant 20 époques avec. En conservant la notation précédente, le coût est calculé par :

$$\mathcal{L}_{\text{CE}}(P|Q) = -\sum_n \sum_k q_n^{(k)} \log p_n^{(k)} \tag{3.7}$$

CHAPITRE 4 RÉSULTATS EXPÉRIMENTAUX

Ce chapitre évalue le gain de performances introduit par les différentes améliorations de l'architecture U-Net décrites précédemment. Puis il compare les performances de segmentation et de classification de la méthode proposée à celles de l'état de l'art et identifie ses limitations.

4.1 Branche basse-résolution seule

L'innovation majeure de l'architecture LRFFCN réside dans l'ajout d'une branche basse résolution opérant sur une image sous-échantillonnée à la résolution de 128×128 pixels. Avant de mesurer le gain de performance sur le réseau complet, vérifions que les caractéristiques extraites par cette branche sont pertinentes pour la classification des vaisseaux.

La validation est menée sur les 20 images de l'ensemble de test de DRIVE (décrite en détail dans la prochaine section) et utilise la prédiction de la couche de classification utilisée lors du pré-entrainement. La précision globale est alors de **93.1%** (toutes classes confondues).

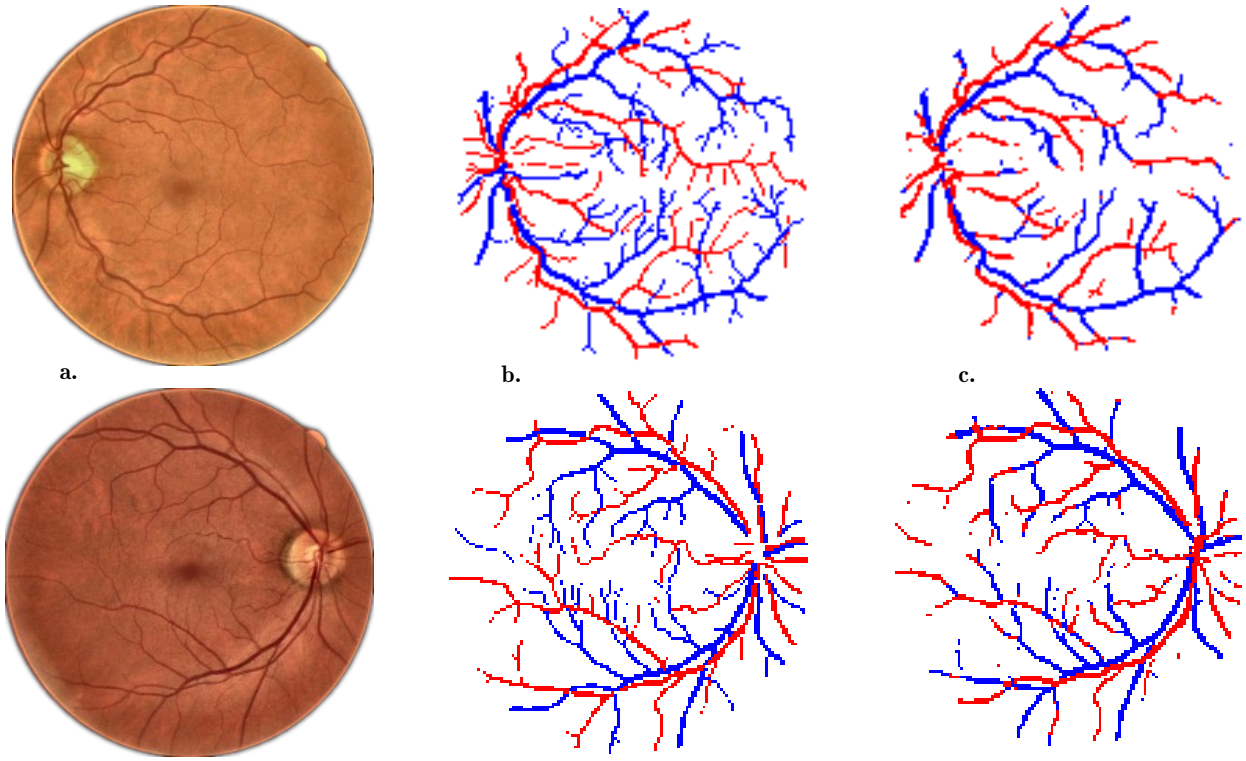

Figure 4.1 Exemples de prédictions de la branche basse résolution. a. Images pré-traitées ; b. Vérité terrain ; c. Carte de segmentation prédite. (Code couleur : *rouge* : artère ; *bleu* : veine).

L'analyse des cartes de segmentation (présenté sur le figure 4.1) indique une bonne segmentation sémantique des vaisseaux moyens et larges malgré quelques incohérences de classification lorsque deux vaisseaux relativement fins sont proches (ces anomalies sont particulièrement visibles sur la seconde image). La majorité des erreurs provient des petits vaisseaux (inférieur à 1 pixels dans l'image source) qui ne sont pas repérés de manière fiable par la branche basse résolution. Mais ces erreurs ne sont pas significatives puisque l'objectif est d'améliorer les performances de classification des vaisseaux moyens et larges.

4.2 Validation sur DRIVE

Lorsqu'ils doivent se comparer entre-eux, la majorité des articles utilise la base de données AV-DRIVE. Cette base de données publiques est composée des 20 images de test de DRIVE annotées par trois cliniciens en cherchant un consensus sur l'annotation (Qureshi et al., 2013). La résolution initiale des images est de 565×594 pixels, mais elles ont été sur-échantillonnées à 2048×2048 pixels.

Le tableau 4.1 montre l'impact de l'ajout de la branche basse résolution et de la couche CRF sur les performances en segmentation et en classification d'un réseau de type U-Net [1]. Comme on pouvait s'y attendre, les 3 architectures obtiennent des performances de segmentation similaires et plutôt hautes, confirmant l'efficacité des réseaux de neurones complètement convolutifs pour cette tache. L'ajout de la branche basse résolution entraine un gain marginal de 0.4% et, contrairement à ce qu'on aurait pu imaginer, le CRF n'améliore pas la segmentation. En effet, la majorité des artéfacts d'acquisition susceptibles de détériorer la qualité de la segmentation vasculaire sont corrigés par le réseau. Mais parfois, la couche CRF qui utilise l'image pré-traitée comme référence, les réintroduits dans les cartes de prédiction.

En ce qui concerne les performances de classifications, l'architecture LRFFCN supplante le U-Net de plus de 3% de précision ! Il semble donc que les caractéristiques contextuelles extraites par la branche basse résolution soient efficacement utilisées par le réseau pour classifier les artères et les veines. On peut confirmer ce résultat en remplaçant le patch d'entrée de la branche basse résolution par du bruit : la précision de classification chute alors de plus de 15%, signe que les informations extraites de ce patch ont une contribution non-négligeable dans la prédiction finale.

De manière plus marginale, l'ajout d'une couche CRF permet d'améliorer la précision moyenne de classification de 0.7% (faiblement significatif au vu de l'écart-type de 3.8%). Par contre elle a un impact bénéfique sur l'équilibre entre la précision de classification des artères et

1. Le réseau U-Net a été entrainé selon la même procédure que le LRFFCN, la seule différence entre leurs deux modèles étant l'absence d'une branche basse résolution.

Tableau 4.1 Performances de segmentation et de classification de l'architecture LRFFCN sur la base de test AV-DRIVE.

	Segmentation	**Classification Artères / Veines**		
Architecture	Précision	Précision	Artères	Veines
U-Net	95.6±0.4%	79.5±3.0%	70.9%	82.6%
LRFFCN	**96.1±0.3%**	82.7±3.3%	79.6%	**85.2%**
LRFFCN + CRF	95.9±0.4%	**83.4±3.8%**	**82.1%**	84.4%

celles des veines : l'écart entre leurs précisions respectives est ainsi réduit de 5.6% à 2.3%.

L'étude comparée des cartes de prédictions (cf. figure 4.2) révèle que la couche CRF corrige une bonne partie des anomalies topologiques et parvient correctement à propager les labels le long des vaisseaux. Cependant, ce sont parfois les erreurs de classification qui sont propagées, ce qui explique le gain de performance mitigé.

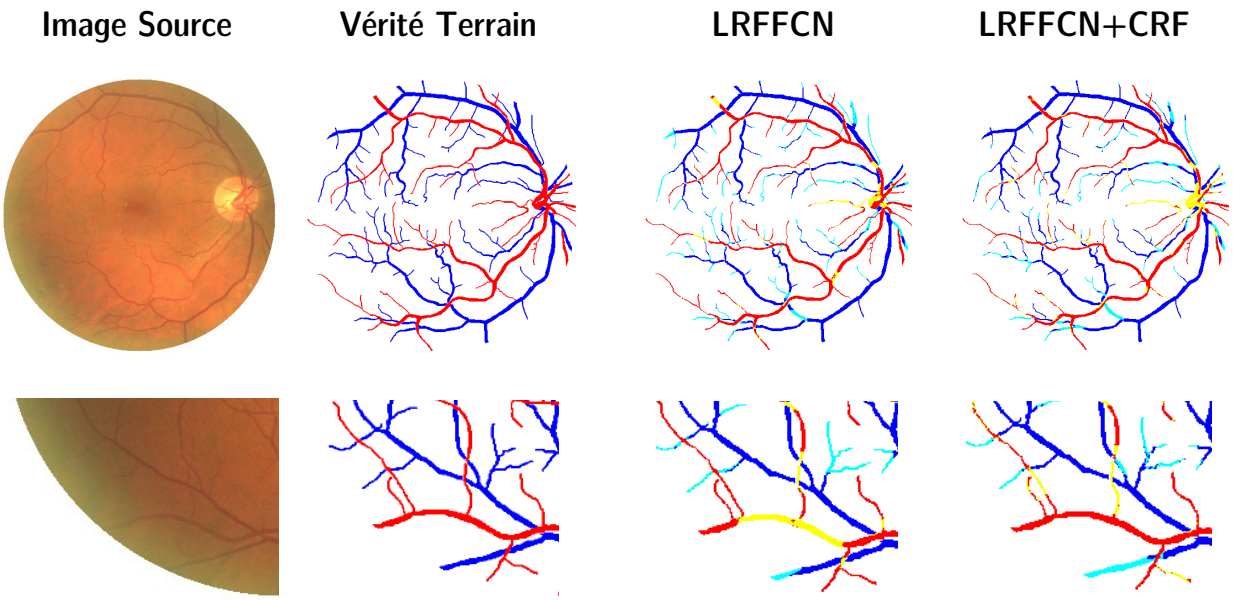

Figure 4.2 Comparaison des performances de l'architecture LRFFCN avec et sans CRFs sur deux image de DRIVE. (Code couleur : *rouge* : vrai artère ; *bleu foncé* : vrai veine ; *bleu clair* : veine classée comme artère ; *jaune* : artère classée comme veine).

4.3 Validation sur MESSIDOR

Le réseau n'a pas été entrainé dans l'optique d'être validé sur DRIVE. En effet, bien qu'elle serve toujours de référence, cette base de données ancienne n'est pas représentative des images de fond d'œil contemporaines. En particulier, leur faible résolution ne permet pas de tirer pleinement profit de l'architecture LRFFCN. De plus DRIVE ne constitue qu'un tiers de la base d'entrainement, les deux tiers restants proviennent de MESSIDOR. 12 nouvelles images de cette base de données ont donc été annotées par nos soins. Cependant, ces images n'ont pas été vérifiées par un clinicien, les performances reportées dans le tableau 4.2 sont donc seulement indicatives. En particulier, le tracé des vaisseaux est de mauvaise qualité et les performances en segmentation ne sont pas significatives.

Même si la qualité des annotations n'est pas comparable avec celle de AV-DRIVE, on peut tout de même constater que la validation MESSIDOR indique une performance de classification supérieur de 1.5% à celle calculée sur DRIVE. La haute résolution des images est donc bien une source importante d'information pour le réseau. De plus, on observe ici aussi que l'architecture LRFFCN offre des performances de classification largement meilleures que l'architecture U-Net naïve, et que l'effet de la couche CRF reste marginal sur la précision de classification, bien qu'il améliore nettement la classification des artères.

Tableau 4.2 Performances de segmentation et de classification de l'architecture LRFFCN sur la base de test MESSIDOR.

	Segmentation	**Classification Artères / Veines**		
Architecture	Précision	Précision	Artères	Veines
U-Net	90.43±1.0%	81.2±5.8%	78.03%	85.1%
LRFFCN	89.95±1.1%	84.2±6.9%	77.39%	**94.3%**
LRFFCN + CRF	**90.89±1.0%**	**84.9±6.1%**	**81.39%**	89.4%

4.4 Comparaison avec l'état de l'art

Les sections précédentes démontrent la pertinence des différentes modifications de l'architecture U-Net pour améliorer ses performances de segmentation et de classification de la vasculature rétinienne. Cette section compare les résultats de l'architecture proposée aux algorithmes de l'état de l'art. Une fois encore, le modèle est évalué sur l'ensemble de test de DRIVE.

4.4.1 Performance de Segmentation

En terme de segmentation, l'architecture LRFFCN obtient de bon résultat. Sa précision de segmentation est supérieure de 1.5% à l'architecture Deep Vessel (Fu et al., 2016) et de 2.1% à celle utilisant l'entrainement adversaire de Lahiri et al. (2017). Comparée à Deep Vessel, l'architecture LRFFCN est bien moins sujette aux faux positifs (sa sensitivité est de 80.8% contre 72.7% pour Deep Vessel, cf tableau 4.3).

Tableau 4.3 Comparaison des performances de segmentations sur DRIVE.

Modèle	Précision	Spécificité	Sensibilité
Mozaffarian et al. (2016)	82.2%	-	-
Lahiri et al. (2017)	94.%	-	-
Deep Vessel, Fu et al. (2016)	94.6%	**97.7%**	72.7%
LRFFCN	**96.1±0.3%**	97.3%	**80.8%**

4.4.2 Performance de Classification

En terme de classification, les résultats restent bien en dessous des meilleurs algorithmes : Estrada et al. (2015a) ont des performances supérieures de 8.3% ! Les écarts de classification sont particulièrement importants sur les petits vaisseaux loin du disque optique, ce qui semble indiquer que le CRF ne parvient pas suffisamment à propager les labels vers les terminaisons vasculaires.

Ces résultats un peu décevants sont à relativiser. D'une part, la faible résolution des images de DRIVE prive le réseau d'informations cruciales pour la classification des vaisseaux les plus fins. D'autre part, la validation a été menée sur toutes les images de l'ensemble de test contrairement à Estrada et al. (2015a) qui furent contraint de ne pas considérer la 11^{e} image de AV-DRIVE (soit 5% de l'ensemble de test) parce qu'une proportion trop importante de vaisseaux étaient déconnectés du disque optique ; ou contrairement au CNN de Welikala et al. (2017) qui atteint 91.27% de précision de classification mais en ne considérant que 10 images de l'ensemble de test (sans expliquer le mode de sélection).

Enfin il faut aussi noter que notre algorithme est particulièrement rapide : une image haute résolution (2048 × 2048 pixels) est traitée en 1.4 secondes (sur GPU : NVidia 1070 Ti), contrairement à l'algorithme fastidieux d'exploration d'arbre de Estrada et al. (2015a) qui a besoin en moyenne de 131 secondes par image pourtant de basse résolution (565 × 594 pixels) !

Tableau 4.4 Comparaison des performances de classification sur DRIVE.

Modèle	Précision	Artères Valides	Veines Valides	Temps
Niemeijer et al. (2011)	80.0%	80.0%	80.0%	-
LRFFCN+CRF	83.4±3.8%	82.1%	84.4%	**1.4 s**
Dashtbozorg et al. (2014)	87.4%	90.0%	84.0%	-
Estrada et al. (2015a)	**91.7±5%**	**91.7%**	**91.7%**	131.3 s

4.5 Limitations de la solution proposée

La principale limitation réside dans le manque de fiabilité lors de la classification de petits vaisseaux. Il est clair que les Champs Aléatoires Conditionnels ne parviennent pas à rectifier les hésitations du réseau de neurones convolutifs, et donnent l'impression d'appliquer des corrections aveugles (l'effet des bonnes corrections étant annulé par les mauvaises). Et en un sens, ils sont bien aveugles : ils n'ont accès à aucune caractéristiques topologiques ou morphologiques pour guider la propagation des labels, excepté la carte de probabilité dont ils sont chargés de la correction... En effet, les simples caractéristiques d'intensités de couleur issues de l'image pré-traitée qui leur sert de référence, sont loin d'être suffisantes pour distinguer les frontières de vaisseaux qui se croisent. Ainsi, si ils parviennent à éliminer les artéfacts de classification au milieu d'un vaisseau, ils ne sont que de peu d'utilité lors de croisements ou de bifurcations.

Le manque de fiabilité de cette propagation de labels est d'autant plus problématique que, hormis la couche CRF, la structure en réseau de neurones complètement convolutif du LRFFCN reste essentiellement une architecture de classificateur unitaire (pixel par pixel). Bien sûr, elle parvient à reconnaitre des structures topologiques à très large échelle, mais elle est parfaitement incapable d'effectuer le suivit d'un vaisseau. Reste à savoir si, comme semble l'indiquer les records de performances de l'état de l'art, seules les méthodes qui intègrent des connaissances cliniques topologiques et qui estiment la structure de l'arbre vasculaire, sont capables de classifier correctement la vasculature rétinienne.

La seconde limitation vient de l'absence de gestion de l'incertitude dans la solution proposée (comme dans toutes celles de l'état de l'art). Dans la base de données DRIVE, certains vaisseaux ne sont pas labellisés comme veines ou artères mais comme une 3^{e} classe incertaine. Il arrive en effet que le type de certains vaisseaux, en particulier les petits vaisseaux émergents du disque optique, soient indistinguables pour un clinicien comme pour un algorithme. Dans ces cas, la gestion d'une information d'incertitude pourrait éviter qu'un algorithme automatique s'appuyant sur ces cartes vasculaires commette des erreurs de diagnostique.

La troisième limitation est intrinsèque au réseau de neurones : leur nature induit un fonctionnement en boite noire dont il est difficile de comprendre le fonctionnement. Ces algorithmes sont capables de généraliser, à partir d'un sous-ensemble de données sur lesquelles ils sont entrainés, des modes implicites qu'ils pourront appliquer avec succès sur la *plupart* des images de l'ensemble initial. Tout le problème réside dans la quantification de ce «plupart». En effet, tant que les raisonnements intrinsèque des réseaux de neurones resteront insondables, il sera impossible de décrire exhaustivement les cas où le réseau se trompera. La seule certitude est que l'augmentation de la base de données d'entrainement diminue le risque d'erreurs.

CHAPITRE 5 CONCLUSION

5.1 Synthèse des travaux

Le travail de recherche décrit dans ce mémoire établi une méthode de segmentation et de classification de la micro-vasculature rétinienne à partir d'images de fond d'œil haute résolution, dans la perspective de l'intégrer à un algorithme de diagnostique automatique de pathologies rétiniennes, cardiovasculaires ou cérébrovasculaires. La majorité des approches traitant de cette question dans l'état de l'art, opèrent la segmentation des vaisseaux séparément de leur classification, puisque la seconde dépend des résultats de la première. Le travail présenté ici propose d'effectuer les deux simultanément. De plus, et contrairement à la grande majorité des méthodes de classifications existantes, il tente d'apprendre la topologie vasculaire de la rétine, plutôt que de l'énoncer en règles dérivées de connaissances cliniques. Pour ce faire, il s'inspire d'un modèle entrainé par apprentissage profond bien connu en traitement d'images médicales : le U-Net. Ce réseau de neurones complètement convolutif est en effet particulièrement performant dans les tâches de segmentations sémantiques. Cependant, les formes fines et allongées des vaisseaux et la ressemblance des petites artérioles et des petites veinules compliquent fortement la tâche du U-Net qui excelle plus dans l'extraction de caractéristiques d'intensités et de textures locales. Ainsi l'application naïve de ce modèle aux images de fond d'œil produit des erreurs de classification locales : au milieu de vaisseaux, ou plus globales : sur des branches entières. Afin d'améliorer les performances sur cette tâche singulière, l'architecture U-Net a été complétée et ses évolutions s'articulent autour de deux objectifs. **1.** Améliorer la classification des vaisseaux larges et moyens en permettant l'apprentissage de caractéristiques topologiques de large échelle. **2.** Améliorer la classification des vaisseaux les plus fins en propageant les labels vers les terminaisons vasculaires.

Pour résoudre le premier problème, une nouvelle architecture de réseau complètement convolutif est développée. Inspirée du U-Net, cette architecture appelée LRFFCN (Réseau Complètement Convolutif à Champ d'Activation Large, Large Receptive Field Fully Convolutional Network) trouve son originalité dans l'ajout d'une branche opérant à une résolution réduite (128×128 pixels au lieu de 2048×2048) est proposé. Cette branche est pré-entrainée sur 130 images et extrait 256 caractéristiques topologiques larges pour un coût en calcul et en mémoire faible. Après ce premier apprentissage, la branche est intégrée à l'architecture du U-Net et ses caractéristiques sont concaténées à celles issues de la branche d'encodage du U-Net. Elles fournissent ainsi des informations contextuelles à l'étage le plus profond du U-Net et contribuent à la prédiction finale en étant interpolées et corrélées dans la branche de décodage.

L'ensemble est alors entrainé à nouveau sur 64 images haute résolution, l'apprentissage du U-Net est alors guidé par les caractéristiques déjà entrainées de la branche basse résolution tandis que les poids de cette branche s'affinent.

Pour solutionner le second problème, l'utilisation des Champs Aléatoires Conditionels (Conditional Random Field, CRF) est proposée. Implémentés comme un couche convolutive récurrente, les CRFs corrèlent, par des noyaux gaussiens, le label de chacun des pixels avec ceux de son voisinage ainsi qu'avec les intensités d'une image de référence (ici l'image pré-traitée). Cette couche permet, une fois intégrée à l'architecture LRFFCNn d'effacer les anomalies topologiques locales et introduit une interdépendance des labels avec leur voisinage respectif. Autrement dit elle rend possible la propagation des labels sur une courte distance.

L'évaluation successive des performances de l'architecture U-Net, LRFFCN et LRFFCN+CRF confirme la pertinence des améliorations proposées. La branche à basse résolution entraine un gain de segmentation de 0.5% mais surtout un amélioration de la précision de classification de 3.8% ! Puis, l'ajout de la couche CRF permet de gagner encore quelques points en classification et équilibre surtout la précision de classification des artères et de veines. Cependant, la comparaison de ces performances avec l'état de l'art offre un bilan plus mitigé. L'algorithme a une meilleure qualité de segmentation des vaisseaux que les algorithmes publiés récemment (un gain de 1.5% est mesuré sur la base de données DRIVE). Mais ses performances de classification sont inférieures de 8.3% à l'algorithme d'exploration exhaustive d'arbre vasculaire de Estrada et al. dont les résultats sont inégalés depuis 2015. Certes les auteurs de cet article durent éliminer une image de la base de test car elle n'était pas suffisamment conforme à leur analyse, certes l'architecture LRFFCN est conçue pour annoter des images plus haute résolution que celle de la vieille base de données publique DRIVE. Mais il semblerait que les CRF ne soient pas l'outil le plus adapté pour propager les labels vers les terminaisons vasculaires puisqu'ils sont incapables de discerner les frontières des vaisseaux aux croisements et aux bifurcations.

Il reste que, la branche basse résolution est un outil efficace pour permettre au U-Net d'accéder à des caractéristiques topologiques larges, et plus généralement, pour augmenter le champs d'activation des réseaux complètement convolutifs contre un coût en ressources réduit. De plus, contrairement à toutes les autres méthodes de classification de la vasculature rétinienne décrites précédemment, l'architecture proposée est la seule dont les performances s'amélioreront à mesure que l'ensemble d'entrainement grossira. Ce travail constitue donc un premier pas prometteur vers la segmentation sémantique du réseau vasculaire rétinien par réseau de neurones convolutifs.

5.2 Améliorations futures

Les limitations évoquées précédemment indiquent plusieurs recommandations, énoncées ici par ordre d'importance :

1. Développer un meilleur algorithme que les CRF pour la propagation des labels au sein de l'arbre vasculaire. Il est notamment possible de modéliser la carte de la vasculature par un maillage de résistances thermiques où les impédances et la température des sources de chaleur sont calculées par le réseau de neurones. On peut alors estimer l'état d'équilibre thermique du système par résolution itérative de l'équation de la chaleur, et la température obtenue sur chaque pixel correspond à sa probabilité d'être une artère ou une veine.
2. Augmenter la taille de l'ensemble d'entrainement. Cette tâche est grandement simplifiée par les performances de notre algorithme, certes imparfait, mais pourtant suffisant pour fournir des cartes de pré-annotations qu'il suffit de corriger (20 minutes plutôt que 2 heures par image).
3. Ajouter une classe «non-distinguable» à la classification artères/veines afin de tenir compte de l'information d'incertitude.
4. Évaluer l'intérêt de l'ajout d'une nouvelle branche intermédiaire opérant à une résolution de 512×512 pixels pour la classification vasculaire.

RÉFÉRENCES

M. D. Abràmoff, M. K. Garvin, et M. Sonka, "Retinal imaging and image analysis." *IEEE reviews in biomedical engineering*, vol. 3, pp. 169–208, 2010. DOI : 10.1109/RBME.2010.2084567

H. Bendaoudi, F. Cheriet, et J. M. P. Langlois, "Memory efficient multi-scale line detector architecture for retinal blood vessel segmentation", dans *2016 Conference on Design and Architectures for Signal and Image Processing (DASIP)*. IEEE, oct 2016. DOI : 10.1109/dasip.2016.7853797

A. Budai, R. Bock, A. Maier, J. Hornegger, et G. Michelson, "Robust vessel segmentation in fundus images", *International Journal of Biomedical Imaging*, vol. 2013, pp. 1–11, 2013. DOI : 10.1155/2013/154860

S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, et M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters." *IEEE transactions on medical imaging*, vol. 8, pp. 263–269, 1989. DOI : 10.1109/42.34715

O. Chutatape, L. Zheng, et S. Krishnan, "Retinal blood vessel detection and tracking by matched gaussian and kalman filters", dans *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*. IEEE, 1998. DOI : 10.1109/iembs.1998.746160

B. Dashtbozorg, A. M. Mendonca, et A. Campilho, "An automatic graph-based approach for artery/vein classification in retinal images", *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1073–1083, mar 2014. DOI : 10.1109/tip.2013.2263809

E. Decencière, G. Cazuguel, X. Zhang, G. Thibault, J.-C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard, R. Danno, D. Elie, P. Massin, Z. Viktor, A. Erginay, B. Laÿ, et A. Chabouis, "Teleophta : Machine learning and image processing methods for teleophthalmology", *IRBM*, vol. 34, no. 2, pp. 196 – 203, 2013, special issue : ANR TECSAN : Technologies for Health and Autonomy. DOI : https://doi.org/10.1016/j.irbm.2013.01.010. En ligne : http://www.sciencedirect.com/science/article/pii/S1959031813000237

Z. X. C. G. L. B. C. B. T. C. . C. B. Decencière, E., "Feedback on a publicly distributed

image database : the messidor database." *Image Analysis and Stereology*, pp. 231–234, 2014.

R. Estrada, M. J. Allingham, P. S. Mettu, S. W. Cousins, C. Tomasi, et S. Farsiu, "Retinal artery-vein classification via topology estimation", *IEEE Transactions on Medical Imaging*, vol. 34, no. 12, pp. 2518–2534, dec 2015. DOI : 10.1109/tmi.2015.2443117

R. Estrada, C. Tomasi, S. C. Schmidler, et S. Farsiu, "Tree topology estimation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 8, pp. 1688–1701, aug 2015. DOI : 10.1109/tpami.2014.2382116

H. Fu, Y. Xu, S. Lin, D. W. Kee Wong, et J. Liu, "Deepvessel : Retinal vessel segmentation via deep learning and conditional random field", dans *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, et W. Wells, éds. Cham : Springer International Publishing, 2016, pp. 132–139.

R. Gunn, "On ophthalmoscopic evidence of general arterial disease", *Trans Ophthalmol Soc UK*, vol. 18, pp. 356–381, 1898.

A. Hoover, V. Kouznetsova, et M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response." *IEEE transactions on medical imaging*, vol. 19, pp. 203–210, Mars 2000. DOI : 10.1109/42.845178

F. Huang, B. Dashtbozorg, et B. M. t. H. Romeny, "Artery/vein classification using reflection features in retina fundus images", *Machine Vision and Applications*, vol. 29, no. 1, pp. 23–34, Jan. 2018. DOI : 10.1007/s00138-017-0867-x. En ligne : https://doi.org/10.1007/s00138-017-0867-x

K. N. K. Noronha, K. Navya, "Support system for the automated detection of hypertensive retinopathy using fundus images", *International Conference on Electronic Design and Signal Processing (ICEDSP)*, pp. 7–11, 2012.

G. Klambauer, T. Unterthiner, A. Mayr, et S. Hochreiter, "Self-normalizing neural networks", dans *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017, pp. 971–980. En ligne : http://papers.nips.cc/paper/6698-self-normalizing-neural-networks.pdf

P. Krähenbühl et V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials", dans *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, et K. Q. Weinberger, éds. Curran Associates, Inc., 2011, pp. 109–117. En ligne : http://papers.nips.cc/paper/

4296-efficient-inference-in-fully-connected-crfs-with-gaussian-edge-potentials.pdf

A. Lahiri, K. Ayush, P. K. Biswas, et P. Mitra, "Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale miscroscopy images", dans *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017, pp. 794–800. DOI : 10.1109/CVPRW.2017.110

W. Luo, Y. Li, R. Urtasun, et R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks", dans *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, et R. Garnett, éds. Curran Associates, Inc., 2016, pp. 4898–4906. En ligne : http://papers.nips.cc/paper/6203-understanding-the-effective-receptive-field-in-deep-convolutional-neural-networks.pdf

K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, et L. Van Gool, "Deep retinal image understanding", dans *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, et W. Wells, éds. Cham : Springer International Publishing, 2016, pp. 140–148.

X. Meng, Y. Yin, G. Yang, Z. Han, et X. Yan, "A framework for retinal vasculature segmentation based on matched filters." *Biomedical engineering online*, vol. 14, p. 94, Oct. 2015. DOI : 10.1186/s12938-015-0089-2

D. Mozaffarian, E. J. Benjamin, A. S. Go, D. K. Arnett, M. J. Blaha, M. Cushman, S. R. Das, et . de Ferranti, "Heart disease and stroke statistics—2016 update", *Circulation*, vol. 133, no. 4, pp. e38–e360, 2016. DOI : 10.1161/CIR.0000000000000350. En ligne : http://circ.ahajournals.org/content/133/4/e38

M. M. K. A. W. D. K. K. N. Iandola, S. Han, "Squeezenet : Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size", *arXiv*, 2016.

U. T. Nguyen, A. Bhuiyan, L. A. Park, et K. Ramamohanarao, "An effective retinal blood vessel segmentation method using multi-scale line detection", *Pattern Recognition*, vol. 46, no. 3, pp. 703 – 715, 2013. DOI : https://doi.org/10.1016/j.patcog.2012.08.009. En ligne : http://www.sciencedirect.com/science/article/pii/S003132031200355X

M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, et M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database", dans *Medical*

Imaging 2004 : Image Processing, J. M. Fitzpatrick et M. Sonka, éds. SPIE, may 2004. DOI : 10.1117/12.535349

M. Niemeijer, X. Xu, A. V. Dumitrescu, P. Gupta, B. van Ginneken, J. C. Folk, et M. D. Abramoff, "Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs." *IEEE transactions on medical imaging*, vol. 30, pp. 1941–1950, Nov. 2011. DOI : 10.1109/TMI.2011.2159619

E. Pellegrini, G. Robertson, T. MacGillivray, J. van Hemert, G. Houston, et E. Trucco, "A graph cut approach to artery/vein classification in ultra-widefield scanning laser ophthalmoscopy", *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 516–526, feb 2018. DOI : 10.1109/tmi.2017.2762963

R. Perfetti, E. Ricci, D. Casali, et G. Costantini, "Cellular neural networks with virtual template expansion for retinal vessel segmentation", *IEEE Transactions on Circuits and Systems II : Express Briefs*, vol. 54, no. 2, pp. 141–145, 2007.

T. A. Qureshi, M. Habib, A. Hunter, et B. Al-Diri, "A manually-labeled, artery/vein classified benchmark for the DRIVE dataset", dans *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. IEEE, jun 2013. DOI : 10.1109/cbms.2013.6627847

E. Ricci et R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification." *IEEE transactions on medical imaging*, vol. 26, pp. 1357–1365, Oct. 2007. DOI : 10.1109/TMI.2007.898551

O. Ronneberger, P.Fischer, et T. Brox, "U-net : Convolutional networks for biomedical image segmentation", dans *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, série LNCS, vol. 9351. Springer, 2015, pp. 234–241, (available on arXiv :1505.04597 [cs.CV]). En ligne : http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a

J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, et B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina." *IEEE transactions on medical imaging*, vol. 23, pp. 501–509, Avr. 2004. DOI : 10.1109/TMI.2004.825627

T.Kauppi, V. Kalesnykiene, J. k. Kamarainen, L. L. I. Sorri, A. Raninen, R. Voutilainen, J. Pietilä, H. Kälviäinen, et H. Uusitalo, "the diaretdb1 diabetic retinopathy database and evaluation protocol", 2007.

D. Toslak, D. Thapa, Y. Chen, M. K. Erol, R. V. Paul Chan, et X. Yao, "Wide-field fundus imaging with trans-palpebral illumination." *Proceedings of SPIE–the International Society for Optical Engineering*, vol. 10045, Jan. 2017. DOI : 10.1117/12.2252491

S. G. Vázquez, B. Cancela, N. Barreira, M. G. Penedo, M. Rodríguez-Blanco, M. P. Seijo, G. C. de Tuero, M. A. Barceló, et M. Saez, "Improving retinal artery and vein classification by means of a minimal path approach", *Machine Vision and Applications*, vol. 24, no. 5, pp. 919–930, jul 2012. DOI : 10.1007/s00138-012-0442-4

R. Welikala, P. Foster, P. Whincup, A. Rudnicka, C. Owen, D. Strachan, et S. Barman, "Automated arteriole and venule classification using deep learning for retinal images from the UK biobank cohort", *Computers in Biology and Medicine*, vol. 90, pp. 23–32, nov 2017. DOI : 10.1016/j.compbiomed.2017.09.005

X. You, Q. Peng, Y. Yuan, Y. ming Cheung, et J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach", *Pattern Recognition*, vol. 44, no. 10-11, pp. 2314–2324, oct 2011. DOI : 10.1016/j.patcog.2011.01.007

D. Yu, H. Wang, P. Chen, et Z. Wei, "Mixed pooling for convolutional neural networks", dans *Rough Sets and Knowledge Technology*, D. Miao, W. Pedrycz, G. Peters, Q. Hu, et R. Wang, éds. Cham : Springer International Publishing, 2014, pp. 364–375.

F. Zana et J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation", *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1010–1019, jul 2001. DOI : 10.1109/83.931095

S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, et P. Torr, "Conditional random fields as recurrent neural networks", dans *International Conference on Computer Vision (ICCV)*, 2015.

ANNEXE A ARCHITECTURE DU RÉSEAU